The Knowledge Engineering Review, Vol. 29:3, 314–344. © Cambridge University Press, 2013
doi:10.1017/S0269888913000180
First published online 3 May 2013

A survey of privacy in multi-agent systems

JOSE M. SUCH, AGUSTÍN ESPINOSA and ANA GARCÍA-FORNES

Departament de Sistemes Informàtics i Computació, Universitat Politècnica de València, Camí de Vera s/n, 46022 València, Spain;
e-mail: jsuch@dsic.upv.es, aespinos@dsic.upv.es, agarcia@dsic.upv.es

Abstract

Privacy has been a concern for humans long before the explosive growth of the Internet. The advances in information technologies have further increased these concerns. This is because the increasing power and sophistication of computer applications offers both tremendous opportunities for individuals, but also significant threats to personal privacy. Autonomous agents and multi-agent systems are examples of the level of sophistication of computer applications. Autonomous agents usually encapsulate personal information describing their principals, and therefore they play a crucial role in preserving privacy. Moreover, autonomous agents themselves can be used to increase the privacy of computer applications by taking advantage of the intrinsic features they provide, such as artificial intelligence, pro-activeness, autonomy, and the like. This article introduces the problem of preserving privacy in computer applications and its relation to autonomous agents and multi-agent systems. It also surveys privacy-related studies in the field of multi-agent systems and identifies open challenges to be addressed by future research.

1 Introduction

Privacy should not be seen as a problem associated only to new technologies (Yao *et al.*, 2007). Indeed, privacy has been a concern long before the emergence of information technologies and the explosive growth of the Internet. There are studies that suggest that privacy is probably as old as the human race itself (Schermer, 2007). Even in primitive societies individuals have always had a desire for privacy (Westin, 1984). This desire for privacy is usually related to the tendency toward territoriality that most animals have. Moreover, the claim of a right for privacy is often related to the instinct of defending against intrusion.

The modern conception of privacy started more than a hundred years ago, with the seminal work of Warren and Brandeis (1890): *The right of privacy*. These two lawyers defined privacy as 'the right to be let alone'. They were pioneers in considering the implications of technology in privacy. Specifically, they were very concerned about the implications of instantaneous photographs and portraits in injuring the feelings of the people in those photographs and portraits. Privacy was later recognized as a fundamental human right by the United Nations Declaration of Human Rights, the International Covenant on Civil and Political Rights, the Charter of Fundamental Rights of the European Union, and many other international treaties (Acquisti *et al.*, 2008).

In the second part of the 20th century, Alan Westin defined privacy as 'the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated' (Westin, 1967). This is what is currently known as the informational self-determination right (Rannenberg *et al.*, 2009). The concept of informational self-determination changed the right to privacy from the right to be let alone to its current incarnation as a means to limit the abuse of personal data (Schermer, 2007). Informational

https://doi.org/10.1017/S0269888913000180 Published online by Cambridge University Press

self-determination represents today's European understanding and regulation of privacy in the context of information and communication technology (EU Directives 95/46/EC, 45/2001/EC, and 2002/58/EC).

Despite all these regulations, as the Internet has no governing or regulating body, privacy breaches are still possible. Nowadays, in the era of global connectivity (everything is interconnected anytime and everywhere) with almost 2 billion world-wide users with connection to the Internet as of 2010[1], privacy is of great concern. In the real world, everyone decides (at least implicitly) what to tell other people about themselves. In the digital world, users have more or less lost effective control over their personal data. Users are therefore exposed to constant personal data collection and processing without even being aware of it (Fischer-Hübner & Hedbom, 2008). Garfinkel (2001) suggests that nowadays users have only one option to preserve their privacy: becoming hermits and not using online social networks, e-commerce sites, etc. Considering the increasing power and sophistication of computer applications that offer many advantages to individuals, becoming a hermit may not really be an option. However, all of these advantages come at a significant loss of privacy (Borking *et al*., 1999). Recent studies show that 90% of users are concerned or very concerned about privacy (Taylor, 2003). Moreover, almost 95% of web users admitted they have declined to provide personal information to web sites at one time or another when asked (Hoffman *et al*., 1999).

Autonomous agents are likely used for personal purposes, so that they usually encapsulate personal information of real people. Consider the case in which agents act on behalf of their principals[2]. In this case, an autonomous agent usually encapsulates personal information describing its principal (Fasli, 2007b), such as preferences, names, and other information. Moreover, agents carry out interactions on behalf of their principals so that they exchange personal information. Agents act on behalf of their principals in agent-mediated e-commerce (Sierra, 2004), as personal assistants (Mitchell *et al*., 1994), in virtual worlds like Second Life[3] (Weitnauer *et al*., 2008), as recommenders (Montaner *et al*., 2003), in agent-mediated knowledge management (Elst *et al*., 2004), in agent-based semantic Web services (Gibbins *et al*., 2004), in distributed problem solving (Wallace & Freuder, 2005), and many other current and future applications. Therefore, they play a crucial role in safeguarding and preserving their principals' privacy. Piolle *et al*. (2007) claim that a great number of researchers in the agent community acknowledge the importance of privacy and believe that more efforts should be made to improve privacy in multi-agent systems.

Surveys such as the one presented in this article are crucial to promote and encourage research and advances in the field of privacy and multi-agent systems. This article introduces the problem of preserving privacy in computer applications and its relation to autonomous agents and multi-agent systems. Therefore, it can be used as a reference for researchers interested in both privacy and multi-agent systems. This article surveys studies that enhance privacy in multi-agent systems as well as studies that enhance privacy through the use of multi-agent systems. It is intended to provide balanced but critical presentations of these studies. These studies are classified according to the type of privacy threats they avoid. Finally, we identify some open challenges in the area of privacy and multi-agent systems.

The remainder of the paper is organized as follows. Section 2 overviews information privacy and its relation to multi-agent systems. Section 3 presents a survey of privacy-preserving mechanisms developed against information collection. Section 4 surveys studies that deal with the avoidance of information processing. Section 5 surveys studies that focus on limiting information dissemination. Section 6 discusses open challenges regarding privacy and multi-agent systems. Finally, section 7 presents some concluding remarks.

[1] http://www.internetworldstats.com/stats.htm

[2] In this paper, we use the terms principal and user indistinctly to refer to the user that the agent is acting on behalf of. Principals are also called agent owners, or simply users in the related literature.

[3] http://secondlife.com/

https://doi.org/10.1017/S0269888913000180 Published online by Cambridge University Press

2 Privacy and agents

There are many studies that treat privacy as a unitary concept with a uniform value, which is unvarying across different situations. However, privacy violations usually involve several types of harmful or problematic activities (Solove, 2002). There are some taxonomies and categorizations that aim to classify these activities (Bostwick, 1976; Kang, 1998; Solove, 2006; Spiekermann & Cranor, 2009). From these taxonomies, Solove (2006) and Spiekermann and Cranor (2009) are commonly accepted as covering most of the privacy aspects. Although we find the taxonomy proposed by Solove (2006) and the one proposed by Spiekermann and Cranor (2009) equally correct and suitable, we use the taxonomy proposed by Solove (2006) because we think it helps the reader to figure out which entity holds the sensitive information at a given moment and which actions the subject of this information can do (or could have done) to protect its privacy.

According to Solove (2006), privacy can be threatened by three main information-related activities: information collection, processing, and dissemination. Information collection refers to the process of gathering and storing data about an individual. Information processing refers to the use or transformation of data that has been already collected. Information dissemination refers to the transfer of collected (and possibly processed) data to other third parties (or making it public knowledge).

Figure 1 depicts a visual scheme that details when information-related activities can be performed in the process of information exchanges among agents. Information collection occurs when agent A communicates personal information about its principal to agent B. In this case, agent B is the one that collects the information. Moreover, although not depicted in the figure for the sake of clarity, a malicious agent could overhear the communications between agent A and agent B and collect information about A. Once agent B has collected information about agent A, it can then process this information. Finally, agent B can disseminate the information it has about agent A (processed or not) to agent C.

The information-related activities described above can represent a chance to breach the privacy of an agent's principal. Examples of possible privacy breaches that can emerge due to these activities are, but not limited to:

- **Secondary use** refers to the use of collected information for purposes different from the purposes for which the information was initially collected and without the data subject's consent for doing so (Solove, 2006). There are potentially infinite types of secondary uses. In the following, we describe some of these possible *secondary uses*:
  - **Profiling**. Hildebrandt and Gutwirth (2008) define profiling as 'the process 'of ''discovering'' patterns in data that can be used to identify or represent a human or non-human subject (individual or group) and/or the application of profiles (sets of correlated data) to individuate and represent an individual subject or to identify a subject as a member of a group (which can be an existing community or a discovered category) and/or the application of profiles to individuate and represent principals or groups'. According to this definition, profiling can be achieved through information collection and processing. One of the most common types of profiling is called buyer profiling in e-commerce environments, in which vendors obtain detailed profiles of their customers and tailor their offers regarding customer's tastes.

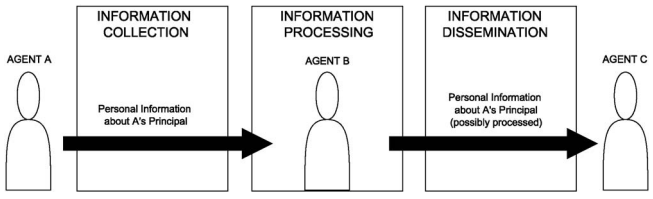


Figure 1 Information-related activities that can threaten privacy

https://doi.org/10.1017/S0269888913000180 Published online by Cambridge University Press

  - **Price discrimination**. Vendors could charge customers different prices for the same good according to the customers' profiles (Odlyzko, 2003), that is, if a vendor knows that some good is of great interest to one customer, the vendor could charge this customer more money for this good than other customers for the same good. For instance, in 2000, Amazon started to charge customers different prices for the same DVD titles (Spiekermann, 2006). When the story became public, Amazon claimed that this was part of a simple price test and discontinued this practice.
  - **Poor judgment**. This is when principals are judged and subsequently treated according to decisions made automatically based on incorrect or partial personal data (Smith & Milberg, 1996). For instance, companies usually divide their potential customers into similar groups based on customers' characteristics (known as customer segmentation). This practice can lead to exclusion of people from services based on potentially distorted judgments (Spiekermann & Cranor, 2009).
- **Identity theft** is 'fraud or another unlawful activity where the identity of an existing person is used as a target or tool without that person's consent' (Koops & Leenes, 2006). For instance, Bilge *et al*. (2009) present how to clone an existing account in an online social network and to establish a friendship connection with the victim in order to obtain information about her/him.
- **Spy agents**. An agent could transfer information about its principal to other third parties without its principal's consent and without its principal being aware of the transfer. For instance, an agent provider that hires or sells agents to principals may design and develop these agents so that they collect information on the principals and their activities (Bygrave, 2001).
- **Unauthorized access**. Sensitive information about principals is transferred on-line even across the Internet and is stored in local and remote machines. Without appropriate protection mechanisms, a potential attacker could easily obtain information about principals without their consent. For instance, an attacker can be listening to transferred information over the network (files, messages, e-mails, etc.) and simply gather the information flowing in the network (Stallings, 2010). This is usually solved by encrypting the information exchanged over a network.
- **Traffic analysis**. Although information exchanged over a network is encrypted, a potential attacker could also gather information about who is communicating with whom. This is because there is information such as the IP address and other where about information of both sender and receiver that is available even if the content of the transferred network packet is encrypted. Thus, this potential attacker could also know how often two individuals communicate to each other and even infer that two individuals are closely related to each other (Korba *et al*., 2002).
- **Unauthorized dissemination or exposure** refers to the transfer of previously collected and possibly processed information to other third parties, which are different from the one that collected (processed) the information, without the consent of the subject of this information. For instance, an agent A collects (and possibly processes) the information that it receives about another agent B. Agent A can transfer information about agent B to another agent C for whatever reason, for example, to receive a monetary compensation. Thus, agent C can perform any of the aforementioned privacy breaches. Moreover, agent C could even make the information about agent A public knowledge, for example, if agent C publishes the information about agent A, for example, in an (online) journal/blog.

All of these privacy breaches may injure the feelings of the principal involved, as pointed out more than a 100 years ago by Warren and Brandeis (1890). Moreover, these privacy breaches can cause other consequences to the principal involved, such as money loss. These privacy breaches could even cause a principal to be summoned by a court, for example, an attacker can steal the identity of a principal and impersonate her/him to carry out unlawful behaviors.

Agents play a crucial role in safeguarding and preserving their principals' privacy (Such, 2011). They usually have a detailed profile of their principals' names, preferences, tastes, location (permanent address, geo-location at a given time), social characteristics (affiliation to groups, friends), roles in organizations and institutions, transactions performed, credit card numbers, and

https://doi.org/10.1017/S0269888913000180 Published online by Cambridge University Press

other personal information. To our knowledge, privacy is seldom considered in the multi-agent systems research field. This leads to agent-based applications that invade principals' privacy, causing concerns about their use and the privacy breaches explained above.

It is crucial for multi-agent systems to consider privacy in order to be of wide use. This can potentially promote principals' trust in agent-based technologies. This trust is needed for principals to be willing to engage with and delegate tasks to agents (Fasli, 2007a). To this aim, studies that enhance privacy in multi-agent systems technologies are needed. Moreover, agent designers and developers also need to be mindful of possible privacy implications when developing agent-based applications (Chopra & White, 2007). This means that agent designers and developers should choose to apply multi-agent systems technologies that preserve privacy, instead of multi-agent systems technologies that are unconcerned about privacy.

Despite having the potential to compromise their principals' privacy, multi-agent systems can also be used to preserve it (Solanas & Martìnez-ballesté, 2009). Multi-agent systems can offer themselves opportunities to enhance privacy beyond what other disciplines in information sciences can do due to their intrinsic features such as intelligence, pro-activeness, autonomy, and the like. According to (Westin, 1967), privacy can also be seen as a 'personal adjustment process' in which individuals balance 'the desire for privacy with the desire for disclosure and communication'.

Humans have different general attitudes toward privacy that influence this adjustment process (Westin, 1967; Ackerman *et al*., 1999; Olson *et al*., 2005): privacy fundamentalists are extremely concerned about privacy and reluctant to disclose personal information; privacy pragmatists are concerned about privacy but less than fundamentalists and they are willing to disclose personal information when some benefit is expected; and finally, privacy unconcerned do not consider privacy loss when disclosing personal information. This view of privacy requires a dynamic management of privacy instead of a static one (Palen & Dourish, 2003). Multi-agent systems can help to support this dynamism, as we will see during this survey.

3 Protection against information collection

As described above, information collection can play a key role in breaching privacy, that is, collected data about a principal can be used to breach her/his privacy. In this section, we describe works in the agent research field that prevent undesired collection of sensitive information. According to Spiekermann and Cranor (2009), information collection involves data transfer and data storage. For the case of agents, this means that information collection involves one agent sending sensitive information to another agent, and that both agents are able to store this sensitive information. Therefore, it is crucial for agents to first decide which information to transfer to which other agent by means of a decision-making mechanism (as described in Section 3.1), and then transfer and store it securely using traditional security mechanisms, such as those that provide confidentiality (as described in Section 3.2).

Another approach for avoiding undesired information collection is the use of third parties. In this case, agents does not send the information directly to the intended destination agents, instead agents provide sensitive information to third parties. These third parties process the information and return the obtained outcomes to the intended destination agents. We describe studies that follow this approach in Section 3.3.

Figure 2 depicts a conceptual map for all of the studied approaches that provide support for protecting against information collection.

3.1 Disclosure decision making

The first important approach to prevent information collection is to decide exactly which information to disclose to other agents. Agents should be able to decide which information to disclose according to their principals' preferences about privacy. This is crucial to prevent undesired information collection. Thus, agents need to incorporate disclosure decision-making

https://doi.org/10.1017/S0269888913000180 Published online by Cambridge University Press

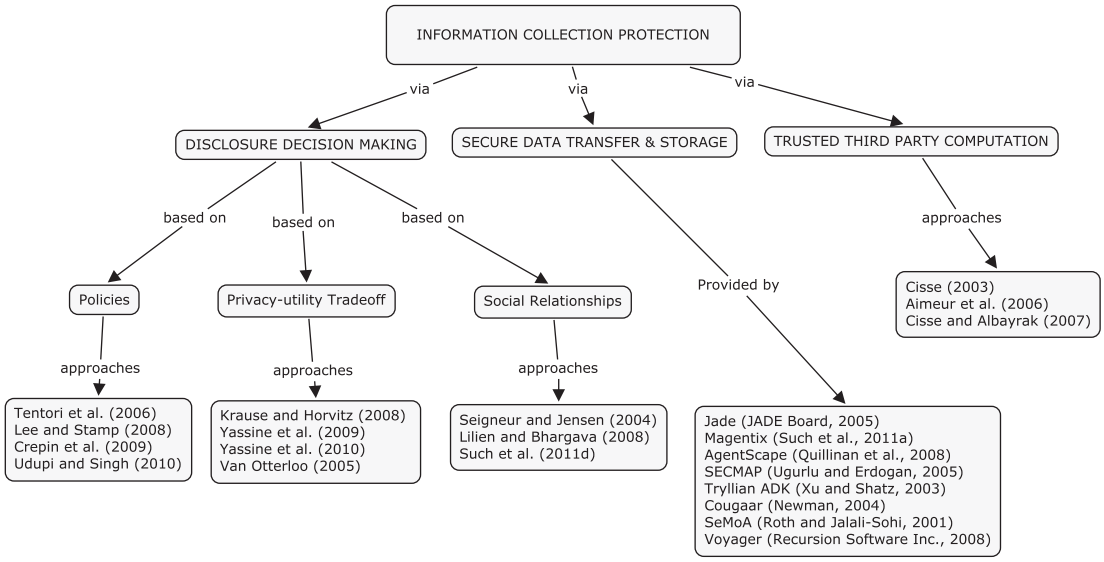


Figure 2 Information Collection Conceptual Map

mechanisms allowing them to decide whether disclosing personal information to other agents is acceptable or not.

3.1.1 Based on policies

One approach for disclosure decision making is based on policies. In this approach, agents usually specify their policies for both disclosing information and the information they want to collect from others. Then, if an agent's policy for disclosing information matches another agent's policy for collecting information from others, the former agent sends the information to the latter.

Tentori *et al.* (2006) presents a privacy-aware agent-based framework that allows agent developers to indicate two privacy-related policies (following an XML schema) per agent: one specifying the privacy policy for information that the agent communicates to others and the other specifying the privacy policy for the information that it requires. There is an agent broker that checks that both policies are compatible. Then, the agent broker monitors and ensures that the information that the two agents exchange complies with the policies. Although it allows the real compliance of privacy policies to be checked, the agent broker becomes a clear performance bottleneck and a single point of failure (SPOF). Moreover, the agent broker itself can be a source of privacy concerns because it knows all the information that two agents communicate to each other.

Lee and Stamp (2008) present an approach based on P3P[4]. The privacy-enhancing agent (PEA) is in charge of automatically retrieving P3P policies of service providers and evaluates whether or not these policies are compliant with its principal's policy. When a principal attempts to access a website, PEA automatically retrieves the website P3P policy and compares it to its principal's preferences. If PEA detects potential privacy violations (i.e. the principal's preferences and the website's P3P policy do not match) or is unable to read the policy of the website, it notifies its principal so that the principal can decide to desist in accessing the website. This approach does not consider that a website may not comply with its announced policy, and, thus, principals' privacy breaches are still possible.

Crépin *et al.* (2009) present an ontology described using Web Ontology Language[5]. Agents can define their policies using this ontology in terms of the hippocratic MAS (Crépin *et al.*, 2008)

[4] The Platform for Privacy Preferences 1.0, `http://www.w3.org/TR/P3P/`

[5] OWL, `http://www.w3.org/tr/owl-features/`

https://doi.org/10.1017/S0269888913000180 Published online by Cambridge University Press

concepts. They differentiate between data provider and data consumer agents. Both of them define their privacy policies according to this ontology. They also propose a protocol by which data consumers request sensitive data from data providers. Data consumers include their policies in the request. If the policy matches the data provider's preferences, the data provider sends the consumer the requested sensitive data. If not, the data provider proposes some modifications to the policy in order to reach an agreement. If the data consumer accepts these adaptations, the data provider sends the requested sensitive data to the data consumer, otherwise the consumer cancels the transaction. Again, they do not consider that data consumers may not comply with the policy they committed to.

Udupi and Singh (2010) present InterPol, a policy language and a framework for interaction in multi-agent referral networks. Policies are logic rules that can be implemented in Prolog. Policies can dynamically adapt to changes in the relationships with other agents. Policies are dynamic in the sense that new predicates can be added to the agent's Knowledge Base. InterPol provides two privacy mechanisms: (i) marking a rule of fact with its visibility (public or private); and (ii) using two privacy-related predicates, servicePrivacyNeed and agentPrivacyTrust, that dynamically manage the privacy decision making. This decision making is based on the privacy that an agent needs for a service and the trust it has in other agents when dealing with privacy issues. However, they do not provide any mechanism for deciding how and when both servicePrivacyNeed and agentPrivacyTrust should be updated.

All of the works presented in this section check that the data exchanged complies with the policies. However, none of them checks that once the data is collected it is treated as stated in the policies. Thus, an agent can disclose personal information to another agent expecting that this second agent will comply with its policy. However, this second agent may not comply with the policy, incurring in possible breaches of privacy.

3.1.2 Based on privacy-utility tradeoffs

There are a great number of people that are willing to trade part of their privacy in exchange for some benefit (64% of US citizens according to Taylor (2003)). They are known as privacy pragmatists, as mentioned above. There are many studies that have focused on providing models of the so-called privacy-utility tradeoff (Lebanon *et al.*, 2006; Krause & Horvitz, 2008). The decision in this case is whether or not a particular privacy-utility tradeoff is acceptable for disclosing information, and then allowing the destination party to collect this information.

The privacy-utility tradeoff is usually modeled as follows. Given a set of personal data attributes A, the utility function of disclosing these attributes $U(A)$, and the privacy cost function of disclosing these attributes $C(A)$, the privacy-utility tradeoff is modeled as $A^* = \arg\max_A U(A) - C(A)$. An example of utility function is the one used by Krause and Horvitz (2008) that measures the reduction of time for performing an online search if some personal data attributes such as the geographical location are given. The privacy cost is usually defined taking into account the sensitivity of the information to be disclosed.

Yassine and Shirmohammadi (2009) present an agent-based architecture that negotiates a reward to be paid to agents' principals by the service providers in return for their disclosures. The data categorization agent is in charge of classifying principals' information into different categories (Yassine *et al.*, 2010). The data categorization agent is able to calculate the privacy cost of the information about its principal considering the categories that this information falls into and the sensitivity for these categories. Principals define the sensitivity for each category and each service provider. The privacy cost is used to calculate an expected reward for disclosing the information. The payoff negotiator agent negotiates a reward for the information disclosed with the service provider, discarding any deal that provides its principal less than her/his expected reward. In this approach, principals must define the expected privacy cost for each category and for each service provider. This can be a burden for principals when considering a large number of service providers.

Another different approach is the one presented by van Otterloo (2005). The author does not focus on information directly disclosed to another party but on the information that can be

https://doi.org/10.1017/S0269888913000180 Published online by Cambridge University Press

collected by observing the strategies an agent follows in a game. The author defines minimal information games as games in which the agent tries to maximize its utility while minimizing the privacy loss. Privacy loss is calculated as the uncertainty (Shannon (1948) entropy) of the strategy that the agent will use. Thus, if an agent uses strategies with high uncertainty, other agents cannot predict their behavior. The author also defines most normal games as games in which the agent tries to maximize its utility while deviating the minimum from the *normal* strategy that other agents will play. In this sense, the agent tries to hide the preferences that differ from the normal behavior of the rest of the agents. The deviation from the normal strategy is calculated as the relative entropy between the agent strategy and the normal strategy. The author of this work does not consider that different actions may have different privacy sensitivity.

The research based on the privacy-utility tradeoff does not consider that there are also other reasons that make people decide whether or not to disclose information about them rather than an increase in utility or a decrease in privacy loss. There are many cases where the direct benefit of disclosing information is not known in advance. The decisions on whether or not to disclose information are based on other reasons in these situations. For instance, the well-known psychological phenomenon called *disclosure reciprocity* (Green *et al.*, 2006) states that one person's disclosure encourages the disclosure of the other person in the interaction, which in turn, encourages more disclosures from the first person.

3.1.3 Based on social relationships

In this section, we describe studies that consider concepts of social relationships when disclosing personal information. In these studies, agents disclose personal information according to the trust and intimacy they have in other agents. In this way, agents decide whether or not to disclose personal information considering trust and intimacy on the one hand and privacy loss on the other hand.

The privacy-trust tradeoff (Seigneur & Jensen, 2004; Lilien & Bhargava, 2008) states that an agent is willing to disclose personal information in order to increase the trust others have in it. Agents are meant to maximize privacy-trust tradeoffs to gain a certain level of trust with the least loss of their privacy. However, the authors of these works associate the different levels of payoff for an agent with the different trust levels that others have in an agent. Then, a trust level is matched with a direct benefit so that an increase in trust results in an increase in utility. As a result, this can be finally modeled as a privacy-utility tradeoff.

There are many cases where the direct benefit of disclosing personal data attributes is not known in advance. This is the case in human relationships, where the disclosure of personal data attributes in fact plays a crucial role in the building of these relationships (Green *et al.*, 2006). These relationships may or may not eventually report a direct benefit for an individual. For instance, a close friend tells you what party he voted for. He may disclose this information without knowing (or expecting) the future gains this may cause in utility. Indeed, this disclosure may not report him any benefit. Moreover, current disclosure decision-making models based on the privacy-utility tradeoff do not consider repeated disclosures and their implications. These implications have been broadly studied in psychology, which has lead to findings regarding how humans disclose personal information in the building of their relationships, such as the well-known *disclosure reciprocity*-phenomenon (Green *et al.*, 2006). This phenomenon is based on the observation that one person's disclosure encourages the disclosure of the other person in the interaction, which in turn, encourages more disclosures from the first person.

Such *et al.* (2011d, 2012a) propose a disclosure decision-making mechanism that considers the disclosure reciprocity phenomenon and that disclosures may not report any benefit (or this benefit may not be known in advance). It is based on intimacy and privacy measures to deal with these situations. Then, agent A may choose to disclose a piece of personal information to agent B so that it maximizes the estimation of the increase in intimacy while at the same time minimizing the privacy loss. Agent A does so expecting that this disclosure would encourage agent B to disclose personal information to agent A. To deal with agents that do not reciprocate the disclosures they

https://doi.org/10.1017/S0269888913000180 Published online by Cambridge University Press

receive, the authors propose a trust model that considers all the disclosures an agent has made to other agent and the reciprocations it received from this agent in order to predict future reciprocations. Moreover, this proposal also considers the reliability of agents when they disclose information about them, that is, how sincere agents are when they disclose information about themselves. The authors of this work assume that personal attributes can be verified and propose a reliability model based on this. However, although most personal attributes can be verifiable directly or indirectly (e.g. tastes can be inferred from what an agent buys), a few others may not be verifiable, for example, the party an agent votes for using an e-voting system.

3.2 Secure data transfer and storage

Once an agent has decided what information to disclose to which agent, this information must be protected from access by any other third party that is different from the agent that the information is intended for. This includes parties from their local computer and network but also different locations, even across the Internet. As stated by Petkovic and Jonker (2007), privacy protection is only possible using secure systems. Security and privacy are often related to each other, but they are not the same (Head & Yuan, 2001). On the one hand, security is the control of information in general (Camp, 1996). Thus, information is secure if the *owner* of the information can control it. On the other hand, information is private if the *subject*of information can control it. Thus, privacy requires security to control access and distribution of information (Garfinkel, 2009).

Agent platforms (APs) provide all the basic infrastructure (for message handling, tracing and monitoring, run-time management, and so on) required to create a MAS (Wooldridge, 2002). There are many APs developed by the agent community—for an overview of current APs and the features they provide refer to Alberola *et al.* (2010). As APs are in charge of executing MAS, they need to be concerned about basic security concepts. However, only a few of them currently take security concerns into account. For instance, Jade (JADE Board, 2005), Magentix (Such *et al.*, 2011a), AgentScape (Quillinan *et al.*, 2008), SECMAP (Ugurlu & Erdogan, 2005), Tryllian ADK (Xu & Shatz, 2003), Cougaar (Newman, 2004), SeMoA (Roth & Jalali-Sohi, 2001), the one presented by Ismail (Ismail, 2008) and Voyager (Recursion Software Inc., 2008) are security-concerned APs.

There are security concepts that are necessary for preserving privacy, such as confidentiality (Gangopadhyay, 2001). All of the above APs use different mechanisms for providing confidentiality. Confidentiality is a security property of a system that ensures the prevention of unauthorized reading of information (Stamp, 2006). This involves both the control of access to information and its distribution. Confidentiality requires authorization mechanisms being in place as well as mechanisms for protecting the transmission of data over a network. They are key mechanisms for avoiding the leak of sensitive information, and, thus, protecting the subject of this information from a privacy breach.

Authorization is the act of determining whether a particular user (or computer system) has the right to carry out a certain activity (Longstaff *et al.*, 1997). The term access control is often used as a synonym for authorization. As an example of a scenario where authorization is necessary for preserving privacy, imagine that two agents A and B are running on the same host and represent two different principals. Agent A may contain a detailed profile about its principal and save this profile in a local file. In this situation, agent B must only be able to access this local file if it is authorized to do so. If agent B succeeds in accessing the file despite not being authorized to do so, this could represent a privacy breach for the principal of agent A.

Security-concerned APs use different authorization mechanisms. These mechanisms allow the specification of rights for agents to carry out activities ranging from traditional access control lists (Jade, Voyager) to other approaches, such as capability-based access control (Ismail, 2008), role-based access control (RBAC; SeMoA, AgentScape, and Tryllian ADK), policy-based access control (Cougaar and SECMAP), and mandatory access control (Magentix). Most APs enforce the access rights that are defined using some of these approaches by means of sandboxing agents.

https://doi.org/10.1017/S0269888913000180 Published online by Cambridge University Press

Confidentiality also implies the protection of transmitted data across a network (Stallings, 2010). In these situations, confidentiality usually means that sensitive information is encrypted into a piece of data so that only parties that are able to decrypt that piece of encrypted data can access the sensitive information. Current security-concerned APs provide confidentiality for the messages exchanged by the agents running on top of them. To this aim, APs use existing secure data transfer technologies such as Kerberos (Neuman *et al.*, 2005), SSL (Frier *et al.*, 1996), and TLS (Dierks & Allen, 1999). These technologies allow the encryption of messages before transferring them and the decryption of messages once they are received. As a result, if an agent A sends a message to an agent B, A is sure that B will be the only one able to read this message.

Without appropriate confidentiality mechanisms, privacy cannot be preserved. All of the APs above provide the needed secure features to secure data transfer and storage. Therefore, they are suitable to prevent undesired collection of information. However, there are also security concepts that can represent by themselves an actual threat for privacy (Petkovic & Jonker, 2007), even though they are mandatory for the system to be secure. For instance, to achieve access control, each entity trying to gain access must first be authenticated so that access rights can be tailored to it (Stallings, 2010). As we will see in Section 4.3, authentication can itself be a threat for privacy. Only some of the security-concerned APs provide mechanisms for preserving-privacy authentication. This should be considered when choosing an AP if there are privacy concerns to be considered.

3.3 Trusted third party computation

Another approach to prevent information collection is based on third parties. Agents provide sensitive information to third parties that process this information and return the outcomes obtained to the intended destination agents. The agent-based information filtering community has developed some proposals that are based on trusted third parties (Cissèe, 2003; Aïmeur *et al.*, 2006; Cissèe & Albayrak, 2007). Information filtering architectures take user profiles and generate personalized information based on them. User profiles usually contain information about preferences, rated items, etc. The resulting systems can be recommender systems, matchmaker systems, or can be a combination of both. The proposals we describe in this section enhance privacy by decoupling the three main parts in an information filtering architecture: users, service providers, and filters.

Aïmeur *et al.* (2006) present a software architecture that they call ALAMBIC. ALAMBIC considers three main parties: users, service providers, and the Still Maker. The Still Maker is a secure platform that generates mobile agents (with a unique public/private key pair) that migrate to service providers. These agents are in charge of filtering the information about users. The code of ALAMBIC agents is encrypted and obfuscated. Moreover, users' profiles are encrypted with the public key of the mobile agents before being transferred to the service provider. As a result, it is very difficult for service providers to obtain more information than the outputs of the filtering process that is carried out inside the mobile agents. However, according to Cissèe and Albayrak (2007), this architecture addresses two aspects inadequately: the protection of the filter against manipulation attempts, and the prevention of collusion between the filter and the provider.

To overcome these aspects, Cissée and Albayrak (2007) propose separating the filter from the service provider. This proposal is based on the use of a trusted AP. Users, service providers, and filter entities (the party that provides filtering functionalities) can deploy agents in the trusted AP. In a nutshell, the information filtering process involves the following steps: (i) the filter entity deploys a temporary filter agent in the trusted AP; (ii) the user entity also deploys an agent, which is called relay agent, in the trusted AP; (iii) the relay agent establishes control of the temporary filter agent (by using mechanisms provided by the trusted AP) and sends the user profile to the temporary filter agent; (iv) the provider profile is propagated from the service provider to the temporary filter agent via the relay agent; (v) the temporary returns the recommendations to the service provider via the relay agent. The authors of this work assume that all providers of APs

https://doi.org/10.1017/S0269888913000180 Published online by Cambridge University Press

are trusted. This assumption may be not valid in truly open multi-agent systems in which there could be untrusted APs.

3.4 Summary of proposals against information collection

To sum up, there are three different approaches that deal with avoiding undesired information collection. First, disclosure decision-making mechanisms (Section 3.1) provide agents with reasoning capabilities for deciding which information to disclose to other agents. Disclosure decision-making mechanisms are currently based on: policies (if the policy of the sender agent matches the policy of the receiver agent, then the sender agent decides to send the corresponding information); privacy-utility tradeoffs (information is only revealed if it is worth it in terms of both the utility and the privacy loss of disclosing this information); and social relationships (information is revealed if the receiver agent is intimate enough or to further increase the intimacy with the receiver agent). Second, secure data transfer and storage mechanisms (Section 3.2) are needed to protect the confidentiality of personal information, so that only authorized parties can access this information. In this way, many APs include mechanisms based on traditional security techniques to provide confidentiality. Third, trusted third party computation mechanisms (Section 3.3) avoid unnecessary information collection by receiver agents. Specifically, agents provide sensitive information to third parties that process this information and return the outcomes obtained to the intended destination agents. Finally, to consult the specific open challenges regarding information collection, we refer the reader to Section 6.

4 Protection against information processing

Information processing refers to the use or transformation of data that has been already collected (Spiekermann & Cranor, 2009). Information processing usually involves various ways of connecting data together and linking it to the individuals to whom it pertains (Solove, 2006). For instance, a vendor could have a complete profile of a customer containing relevant data collected from the purchases made by the customer's agent. The vendor can then use information filtering techniques to obtain detailed information on the customer's tastes. Then, the vendor can infer which goods the customer is more willing to acquire and offer them in advance through personalized advertising. Moreover, the vendor could even incur in price discrimination practices, that is, the vendor could charge different prices to different customers depending on the desire that the customer has to acquire a product according to their tastes.

Most of the work for protecting against the processing of information already collected is based on the principle of data minimization. Data minimization states that disclosed personal data should preserve as much unlinkability as possible (Pfitzmann & Hansen, 2010). This is a way to reduce the probability of different pieces of data being connected to each other and linked to an individual. Therefore, privacy threats are reduced while still allowing information to be collected.

Spiekermann and Cranor (2009) state that 'Identifiability can be defined as the degree to which (personal) data can be directly linked to an individual'. The degree of privacy of a system is inversely related to the degree of user data identifiability (Pfitzmann & Hansen, 2010). The more identifiable data that exists about a person, the less that person is able to control access to information about herself/himself, and the greater the privacy risks. Identifiability ranges from completely identified to anonymous. Throughout this section, we survey different studies in MAS that prevent information processing through minimizing the collection of identifiable data.

Figure 3 depicts a conceptual map for all of the studied approaches that provide support for protecting against information processing.

4.1 Anonymity

Anonymity is the maximum degree of privacy, so it plays a crucial role in preserving privacy in agent technologies (Brazier *et al.*, 2004). The main property of anonymity is that collected data cannot later be attributed to a specific individual. Anonymity is commonly defined in terms of a

https://doi.org/10.1017/S0269888913000180 Published online by Cambridge University Press

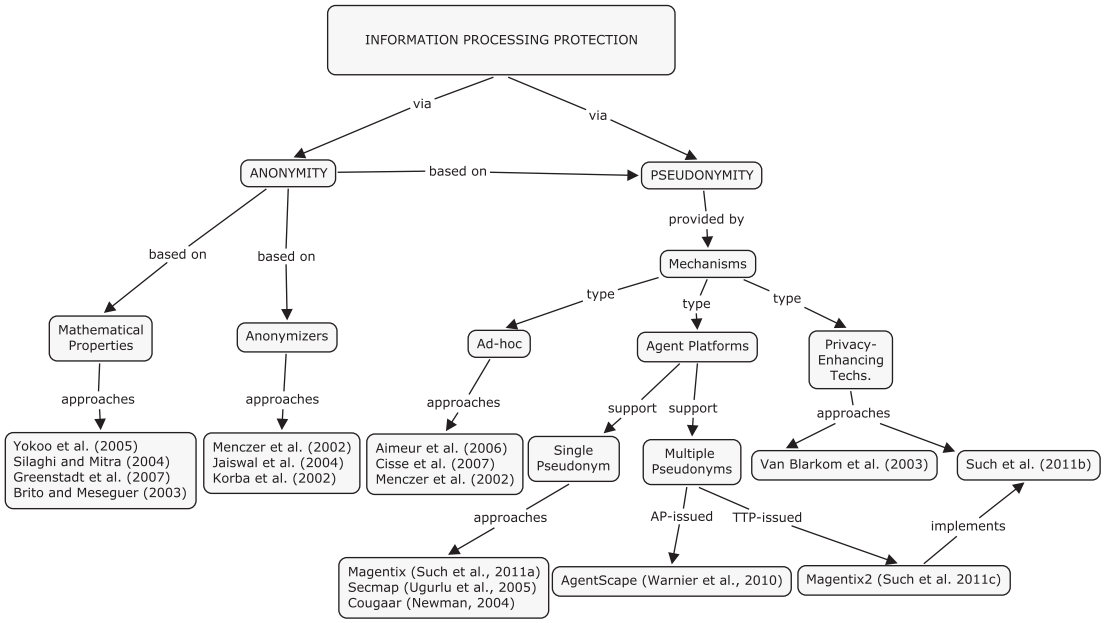


Figure 3 Information Processing Conceptual Map

possible attacker. Pfitzmann and Hansen (2010) define anonymity as 'Anonymity of a subject from an attacker's perspective means that the attacker cannot sufficiently identify the subject within a set of subjects, the anonymity set.'

Many anonymity systems can be modeled in terms of unlinkability (Diaz, 2006). Pfitzmann and Hansen (2010) define unlinkability as follows: 'Unlinkability of two or more items of interest (IOIs, e.g., subjects, messages, actions, etc.) from an attacker's perspective means that within the system (comprising these and possibly other items), the attacker cannot sufficiently distinguish whether these IOIs are related or not'. Anonymity can be achieved when a given IOI cannot be linked to a given subject. For instance, the sender of a message is anonymous if the message cannot be linked to a particular subject from a set of subjects that may have potentially sent the message (the anonymity set).

4.1.1 Anonymity in multi-agent problem solving

The agent community has developed algorithms that aim at preserving anonymity in multi-agent problem solving, including both distributed constraint satisfaction (DisCSP) and distributed constraint optimization (DCOP). In these problems, agents need to share information in order to solve a problem of mutual interest. The major concern in DisCSP and DCOP algorithms is that they usually leak information that can be exploited by some agents to infer private information of other agents (Greenstadt *et al.*, 2006). The anonymity set here is the set of agents that share information. The main aim of these protocols is that shared information cannot be linked to the corresponding agent. A typical application is that of meeting scheduling in which agents arrange meetings according to their principals' schedules. Private information in these problems usually refer to information about: (i) agent preferences (domain privacy), that is, whether an agent can attend a meeting in each time slot in DisCSP or the utility valuations for each agent for each time slot in DCOP; and (ii) the assignment for each agent once a final solution is reached as well as partial solutions during the solving process (assignment privacy).

Yokoo *et al.* (2005) and Silaghi and Mitra (2004) present secure DisCSP algorithms based on multiparty computation. Multi-party techniques compute general functions with secret inputs. Therefore, these techniques allow the collection of information in a way that cannot be linked back to the agents. Theoretical proofs show that these secure DisCSP algorithms do not leak

https://doi.org/10.1017/S0269888913000180 Published online by Cambridge University Press

private information, that is, there is no chance for either an agent or an external entity to link variable assignment to the agents taking part in the problem-solving process. However, these approaches have a high computational cost. DisCSP algorithms require many comparisons and these protocols require exponentiation operations for each comparison. Therefore, these protocols should be used when privacy concerns are very high.

There are other computationally cheaper approaches, such as (Greenstadt *et al.*, 2007) and (Brito & Meseguer, 2003). However, these approaches still leak information. These works try to reduce privacy loss of existing DisCSP/DCOP algorithms. Metrics based on the Valuations of Possible States (Maheswaran *et al.*, 2006) framework are usually considered to quantify the reduction in privacy loss. Greenstadt *et al.* (2007) present the DPOP with Secret Sharing (SSDPOP) algorithm, which is an extension of DPOP based on the efficient cryptographic technique of secret sharing. Agents use secret sharing to send aggregate values, and, thus, they do not reveal their individual valuations. Brito and Meseguer (2003) present the Distributed Forward Checking (DisFC) algorithm, which is an approach without using cryptography. In DisFC, agents exchange enough information to reach a global consistent solution without making their own assignment public. To this aim, DisFC sends filtered domains (agent preferences) to other agents and replaces their own value by a sequence number.

4.1.2 Anonymizers

There are technologies developed outside the agent community called anonymizers. These anonymizers can be used to obtain communication anonymity. They hide the IP address and other where about information from the messages they receive and forward these messages (Menczer *et al.*, 2002). If an agent sends a message to another agent using these anonymizers, the receiving agent is not able to identify the sender from the potential senders (the anonymity set in this case). Chaum (1981) first introduced MIX-networks as a means to counteract traffic analysis. A MIX-network is composed of a set of MIX nodes. Each MIX node receives a number of messages, modifies them (using some cryptographic transformation), and sends them randomly. Moreover, each MIX node in the network knows only the previous and next node in a received message's route. Therefore, an external observer is not able to correlate incoming and outcoming messages.

Onion routing (Goldschlag *et al.*, 1999) is based on Chaum's MIX-networks. While MIX nodes could store messages for an indefinite amount of time while waiting to receive an adequate number of messages to mix together, an onion router is designed to pass information with low latency. However, large message traffic is vital to strengthen anonymity in onion router networks. An example of implementation of onion routing is Tor (Dingledine *et al.*, 2004).

Anonymizers also prevent an external observer from inferring a possible relationship between the sender and the receiver of a message (known as traffic analysis). Although authorization and confidentiality are ensured by secure APs (explained in Section 3.2), a potential attacker could also gather information about who is communicating with whom. This is because the IP of both sender and receiver can be known even though the content of the message exchanged is encrypted. Moreover, this potential attacker could also know how often two agents communicate to each other, and then infer relationship patterns between the two agents.

There are some agent architectures and implementations that use anonymizers (Menczer *et al.*, 2002; Jaiswal *et al.*, 2004). IntelliShopper (Menczer *et al.*, 2002) is an intelligent shopping agent that aids customers who are shopping for a product on an e-commerce site. This agent is in charge of monitoring e-commerce sites to notify the customer about updates related to products she/he is interested in. To this aim, the agent is able to collect information about the customer's activities at an e-commerce site to determine interesting products and decisions about whether or not she/he buys the products. Customers connect to IntelliShopper through an anonymizer so that IntelliShopper cannot know about the IP and other where about information of the requests it receives. The MAGNET architecture (Jaiswal *et al.*, 2004) provides support for auction-based business-to-business marketplaces. MAGNET agents participate in auctions that are reverse, that is, contracting agents present

https://doi.org/10.1017/S0269888913000180 Published online by Cambridge University Press

call for bids to supplier agents. MAGNET uses an anonymizer between the market and the bidders. This is intended to reduce market–supplier collusion by making the supplier's bids unlinkable until the end of auction.

There are also anonymizers specially developed for APs. Korba *et al.* (2002) present an alternate Onion Routing approach for Jade. Each Jade AP has several onion agents that provide an anonymous data forwarding service, and at least one onion monitor agent that keeps track of the location of all the other onion agents in the system. Onion monitor agents exchange information in order to maintain a valid topology of the complete onion agent network. The main drawback of this approach is that agents in Jade do not communicate directly with each other; instead, it is the *container* where agents live that finally sends the message over the network to the container where the recipient agent lives. Therefore, an external observer could track the path of a message through the containers and infer possible relations among agents living in these containers. The lower the number of agents in a container, the higher the probability for an external observer to link the message to one particular agent (as sender or receiver). Moreover, the Jade AP itself could monitor the path of a message from one agent to another. A possible solution for this may be that containers connect to each other through the general purpose anonymizers presented above.

4.2 Pseudonymity

Pseudonymity (Chaum, 1985) is the use of pseudonyms as identifiers. A pseudonym is an identifier of a subject other than one of the subject's real names (Pfitzmann & Hansen, 2010). Pseudonyms have been broadly used by human beings in the real world. For instance, in the 19th century when writing was a male-dominated profession, some female writers used male names for their writings. Nowadays, in the digital world, there are a great number of pseudonyms such as usernames, nicknames, e-mail addresses, sequence numbers, public keys, etc.

The most important trait of pseudonymity is that it comprises all degrees of identifiability of a subject (from identified to anonymous) depending on the nature of the pseudonyms being used. Complete identification is when the linking between a pseudonym and its holder is publicly known. Anonymity can be achieved by using a different pseudonym for each different transaction known as transaction pseudonyms (Chaum, 1985), unless the information contained in these transactions establishes linkability (Bhargav-Spantzel *et al.*, 2007).

4.2.1 Ad hoc mechanisms

There are some agent-based approaches that implement *ad hoc* mechanisms for implementing pseudonymity. Some of these approaches have been proposed in the agent-based information filtering domain (Aïmeur *et al.*, 2006; Cissée & Albayrak, 2007). Cissée and Albayrak (2007) provide an approach based on transaction pseudonyms. They aim at providing anonymity, that is, the recommendations must not be linkable to the identity of the principal if the principal wants these recommendations to be unlinkable to himself. They propose that principals use a different agent each time they ask for a recommendation. Aïmeur *et al.* (2006) provide a completely different approach; the principal 'identifies' herself/himself when she/he communicates with the service provider by using a pseudonym. Thus, the service provider can build a profile to better aid recommendations but without establishing linkability to the identity of the agent's principal. However, none of these approaches consider that the principal may want some recommendations to be unlinkable while others be linkable; instead they provide the use of either only one pseudonym or a different pseudonym for each recommendation.

Other approaches to pseudonymity come from the agent-based e-commerce domain (Menczer *et al.*, 2002). Users connect to the IntelliShopper agent using a pseudonym to avoid the link between the profiles that IntelliShopper has about customers and their real identity. Moreover, users can use different pseudonyms for IntelliShopper to have separate profiles for separate

https://doi.org/10.1017/S0269888913000180 Published online by Cambridge University Press

activities. Therefore, users can decide whether or not to use a new pseudonym in each transaction, instead of forcing the same pseudonym for all transactions or a different pseudonym for each transaction (as in the approaches described in the above paragraph). However, the authors of this work leave the user with the responsibility of creating their pseudonyms and they do not provide any pseudonym management facility.

4.2.2 Support from APs

Another approach for providing general support for pseudonymity for agent technologies instead of *ad hoc* solutions is to provide this support from APs. Thus, this support aids agent developers to use pseudonymity without having to implement their own solutions. However, only a few of the APs explained in Section 3 implement some kind of support for pseudonymity. Magentix (Such *et al*., 2011a), Secmap (Ugurlu & Erdogan, 2005), AgentScape (Quillinan *et al*., 2008) and Cougaar (Newman, 2004) assign a unique identity for each agent that it can use to authenticate itself to other agents. Using this identity, agents can act pseudonymously, that is, agents can act on behalf of their principal without using the identity of their principal. However, agents cannot hold more than one pseudonym, that is, principals should use a different agent each time they want to use a different pseudonym (similarly to what is proposed by Cissée and Albayrak (2007) explained above).

Warnier and Brazier (2010) also present a mechanism for the AgentScape AP that offers pseudonymity by means of what they call *handlers*. Handlers are pseudonyms that agents can use to send/receive messages to/from other agents. At will, agents can request the AP for new handlers. Moreover, the AP is the only one that knows the association between handlers and GUIDs (global unique identities of the agents). An agent can also obtain anonymity by simply using a different handler for each transaction (transaction pseudonyms). AgentScape also offers an automatic anonymity service. Agents can send messages anonymously without having to manage pseudonyms. This service is provided by agents called *anonymizers*[6]. When an agent wants to send a message anonymously, this message is redirected to an anonymizer. Then, this anonymizer is in charge of removing the original handler of the sender from the message, replacing it with another (possibly new) handler, and sending the message to the intended recipient. If the intended recipient replies, this reply is forwarded to the sender of the original message.

The original sender of the message must notify when a transaction ends. For each new transaction, the anonymizer generates a new handler.

APs that provide support for pseudonymity (e.g. by providing APIs to create and manage pseudonyms) do not consider that pseudonyms can be issued by external third parties. That is, APs themselves are in charge of issuing the pseudonyms. Thus, the AP itself (and the anonymizer agents for the case of AgentScape) must be trusted. This is because the AP knows the relation of pseudonyms to each other and to the principal involved. This usually implies that the organization or company that hosts the specific system (e.g. eBay in the case of an e-marketplace) knows the association of pseudonyms to each other and to principals. Therefore, this organization or company can collect and process information about the principals that run their agents on the system.

Other more general approaches have been proposed to provide pseudonymity to agent technologies (van Blarkom *et al*., 2003; Such *et al*., 2011b). Both approaches propose the integration of Privacy-Enhancing Technologies (PETs; Senicar *et al*., 2003) into agent technologies. PETs can be defined as 'a system of ICT measures protecting informational privacy by eliminating or minimising personal data thereby preventing unnecessary or unwanted processing of personal data, without the loss of the functionality of the information system' (van Blarkom *et al*., 2003).

Van Blarkom *et al*. (2003) propose the use of Identity Protectors. Identity Protectors are PETs that are in charge of converting the identity of the principal involved (the person whose

[6] Note that these anonymizers are not the same as the ones presented in Section 4.1.2.

https://doi.org/10.1017/S0269888913000180 Published online by Cambridge University Press

data are being processed) into one or more pseudonyms. They propose that the Identity Protector is placed either between the principal and the agent or between the agent and the environment. The Identity Protector in an information system can take several forms: (i) a separate function implemented in the data system; (ii) a separate data system supervized by the individual; (iii) a data system supervized by a trusted third party. They present in which places of a specific agent architecture an Identity Protector can be placed. However, they do not provide any specific design or implementation of an Identity Protector and the integration of it into an agent architecture.

Such *et al.* (2011b) present a proposal for an Identity Protector based on trusted third parties. Moreover, this proposal has been integrated into the Magentix2 AP (Such *et al.*, 2011c). Specifically, this proposal builds on modern Privacy-Enhancing Identity Management Systems (PE-IMS) (Hansen *et al.*, 2004). PEIMS are PETs that support the management of pseudonyms to control the nature and amount of personal information disclosed. These systems are aimed at first providing the controlled pseudonymity of the principals; and second, the reliability of the principals. Controlled pseudonymity implies unlinkability between the pseudonym and the real identity of the principal behind the partial identity. Controlled pseudonymity also implies that the pseudonyms of the same principal are unlinkable if they are used in different contexts. The reliability of the users implies that, at first, there is unlinkability between pseudonyms and the real identity of the entities behind them, but under special circumstances the issuer of the pseudonym can make a pseudonym and the real identity of an entity linkable. The authors of this approach assume an identity infrastructure to be used as a trusted third party. This assumption can be made in networks such as the Internet, as contemplated by standards such as the Identity meta system interoperability standard[7]. However, this assumption may not hold in environments with very scarce resources such as sensor networks.

4.3 Implications in security, trust, and reputation

4.3.1 Implications in security

As stated in Section 3, security plays a crucial role in preventing undesired information collection. However, there are also security concepts themselves that can represent an actual threat to privacy (Petkovic & Jonker, 2007) even though they are mandatory for the system to be secure. These security concepts include authentication and accountability. Minimizing data identifiability may affect authentication and accountability if specific countermeasures are not considered.

Authentication binds a principal to a digital representation of her/his identity (Bishop, 2002). To authenticate something on the Internet is to verify that its identity is as claimed (Jøsang *et al.*, 2001). In the case of a message, the function of authentication is to assure the recipient that the message is from the source that it claims to be from (Stallings, 2010). For instance, if an agent A sends a message to an agent B, B should be able to *authenticate* A as the sender of the message. Authentication of the entities existing in an AP is the basis for confidentiality (Such *et al.*, 2011a), explained in Section 3. All of the APs that provide some support for pseudonymity (Magentix (Such *et al.*, 2011a), Secmap (Ugurlu & Erdogan, 2005), AgentScape (Quillinan *et al.*, 2008) and Cougaar (Newman, 2004)) support authentication based on pseudonyms.

The other security concept that has a direct impact on privacy (and vice versa) is accountability. Accountability refers to the ability of holding entities responsible for their actions (Bhargav-Spantzel *et al.*, 2007). Accountability helps to promote trust in the system. This is because if an agent misbehaves and there are no accountability consequences, there is a sense of impunity that could even encourage abuse. This trust is crucial for systems in which users can be seriously damaged by losing money, such as agent-based e-commerce (Fasli, 2007b).

Accountability usually requires an unambiguous identification of the principal involved (Bishop, 2002). Then, this principal can be held liable for their acts. For instance, a customer agent

[7] http://docs.oasis-open.org/imi/identity/v1.0/identity.html

https://doi.org/10.1017/S0269888913000180 Published online by Cambridge University Press

pays a vendor agent for a good. The agent vendor commits to shipping the good to the customer agent's principal. In the event that the customer agent's principal does not receive the good, the vendor's principal[8] may be held liable for this. Although determining exactly who should be held liable for this depends on the applicable laws in the specific country, it usually requires the identification of the vendor's principal. Then, the vendor agent's principal can be sued for fraud.

Pseudonyms can be utilized to implement accountability (Hansen *et al*., 2004). AgentScape and Magentix keep track of the association between principals and pseudonyms. Therefore, these two APs can disclose the principal behind the pseudonym, removing pseudonymity and producing identity and accountability as a result. The main drawback of this approach is that the AP itself (including the anonymizer agents for the case of AgentScape) must be trusted. This is because the AP knows the relation of pseudonyms to each other and to the principal involved. Although this is needed to ensure accountability (agent principals can still be held liable for their agents behavior even when pseudonyms are used), this usually implies that the organization or company that hosts the specific marketplace (e.g. eBay) knows the association of pseudonyms to each other and to principals. Therefore, this organization or company can collect and process information about the principals that run their agents in the marketplace.

Other approaches such as Such *et al*. (2011b) allow APs and agents to use pseudonyms generated by trusted third parties that do not participate in the specific marketplace. Therefore, the AP itself does not know the association between pseudonyms and their principals. These third parties may disclose this information only if required by a court. However, this does not prevent these third parties and the AP (or an agent) from colluding to finally obtain this information.

4.3.2 Implications in trust and reputation

In a multi-agent system, agents usually need to assess trust toward other agents as well as their reputation. To this aim, the agent community has developed a vast number of trust and reputation models (Ramchurn *et al*., 2004; Sabater & Sierra, 2005). An agent can build a reputation by using the same pseudonym more than once. In the same way, an agent can be trusted by a transaction partner by using the same pseudonym for different transactions. Current trust and reputation models are usually based on the assumption that pseudonyms are long lived, so that ratings about a particular entity from the past are related to the same entity in the future. However, when these models are actually used in real domains, this assumption is no longer valid. For instance, an agent that has a low reputation due to its cheating behavior may be really interested in changing its identity and restarting its reputation from scratch. This is what Jøsang *et al*. (2007) called the *change of identities* problem. This problem has also been identified by other researchers under different names, for example, *whitewashing* (Carrara & Hogben, 2007).

Kerr and Cohen (2009) also point out the fact that entities could create new accounts (identity in the system) at will, not only after abandoning their previous identity but also holding multiple identities at once. This is known as the *sybil* attack (Jøsang & Golbeck, 2009). An example of this attack could be an agent that holds multiple identities in a marketplace and attempts to sell the same product through each of them, increasing the probability of being chosen by a potential buyer.

These vulnerabilities can be more or less harmful depending on the final domain of the application. However, these vulnerabilities should, at least, be considered in domains in which trust and reputation play a crucial role. For instance, in e-marketplaces, these vulnerabilities can cause users to be seriously damaged by losing money. Another example can be a social network like Last.fm[9] in which users can recommend music to each other. A user who always fails to

[8] Software entities (intelligent agents, virtual organizations, etc.) cannot have real identities because, until now, they could not be held liable for their acts in front of the law. However, this may change in the future if they finally achieve some kind of legal personhood, as suggested by Chopra and White (2004) and Balke and Eymann (2008). In this case, software entities may be provided with legal personhood to be (partially) held liable for their acts. The point is that according to the law, someone must be liable for frauds like this.

[9] Last.fm, `http://www.last.fm`

https://doi.org/10.1017/S0269888913000180 Published online by Cambridge University Press

recommend good music to other users may gain a very bad reputation. If this user creates a new account in Last.fm (a new identity in Last.fm) her/his reputation starts from scratch, and is able to keep on recommending bad music. Users may be really bothered by these recommendations and move to other social networks. In this case, the one seriously damaged is the social network itself by losing users.

A possible solution for these vulnerabilities is the use of *once-in-a-lifetime* pseudonyms (Friedman & Resnick, 1998). Agents can only hold one *once-in-a-lifetime* pseudonym in each marketplace. Therefore, they cannot get rid of the trust other agents have in them as well as the reputation they earned in the multi-agent system. A model for agent identity management based *once-in-a-lifetime* pseudonyms has been proposed in Such *et al.* (2011b). This model considers two kinds of pseudonyms: permanent and regular. Agents can only hold one permanent pseudonym in a given system. Regular pseudonyms do not pose any limitation. Although both kinds of pseudonyms enable trust and reputation relationships, only permanent pseudonyms guarantee that identity-related vulnerabilities are avoided. Then, agents will choose to establish trust and reputation through permanent pseudonyms if they want to avoid identity-related vulnerabilities. If they want to avoid information processing, they will use as many regular pseudonyms as needed to achieve their desired privacy level (they reach the maximum level by using a different pseudonym for each different transaction). However, this model needs the existence of trusted third parties called Identity Providers to issue and verify pseudonyms. While this may not be a difficulty in networks such as the Internet, this may not be appropriate in environments with very scarce resources such as sensor networks.

There are also other solutions for identity-related vulnerabilities of trust and reputation models that can be used when trusted third parties cannot be assumed (Hoffman *et al.*, 2009). Yu *et al.* (2006) present an approach based on social networks represented as a graph in which nodes represent pseudonyms and edges represent human-established trust relationships among them in the real world. They claim that malicious users can create many pseudonyms but few trust relationships. They exploit this property to bind the number of pseudonyms to be considered for trust and reputation. However, this approach is not appropriate for open MAS in which agents act on behalf of principals that may not be known in the real world. (Cheng & Friedman, 2005) have demonstrated several conditions using graph theory that must be satisfied when calculating reputation in order for reputation models to be resilient to sybil attacks. This approach needs a particular and specific way to calculate ratings about an individual. Thus, this approach cannot be applied to trust and reputation models that follow other approaches for managing trust and reputation ratings.

4.4 Summary of proposals against information processing

Over the course of this section, we have detailed two different (but at the same time related) approaches to protect against information processing. The first one is based on anonymity (Section 4.1). Specifically, we have presented approaches that provide anonymity in multi-agent problem solving, in which agents need to share information in order to solve a problem of mutual interest but they want to avoid leaking information that can be exploited by some agents to infer private information. We have also shown technologies that support the anonymization of communications over the Internet by means of what is known as anonymizers. The second approach is based on pseudonymity (Section 4.2). We have shown several *ad hoc* approaches that provide limited pseudonym support. Moreover, we have also described complete pseudonym management systems that some APs provide and that facilitate the development of agents that make extensive use of pseudonyms (which can also lead to achieve anonymity if, for instance, an agent uses a different pseudonym for each new communication). We have also shown in Section 4.3 the interplay between anonymity and pseudonymity on the one hand, and security, trust and reputation on the other hand. We have concluded that anonymity and pseudonymity can have a direct impact on security, trust and reputation, which also play a crucial role to preserve privacy.

https://doi.org/10.1017/S0269888913000180 Published online by Cambridge University Press

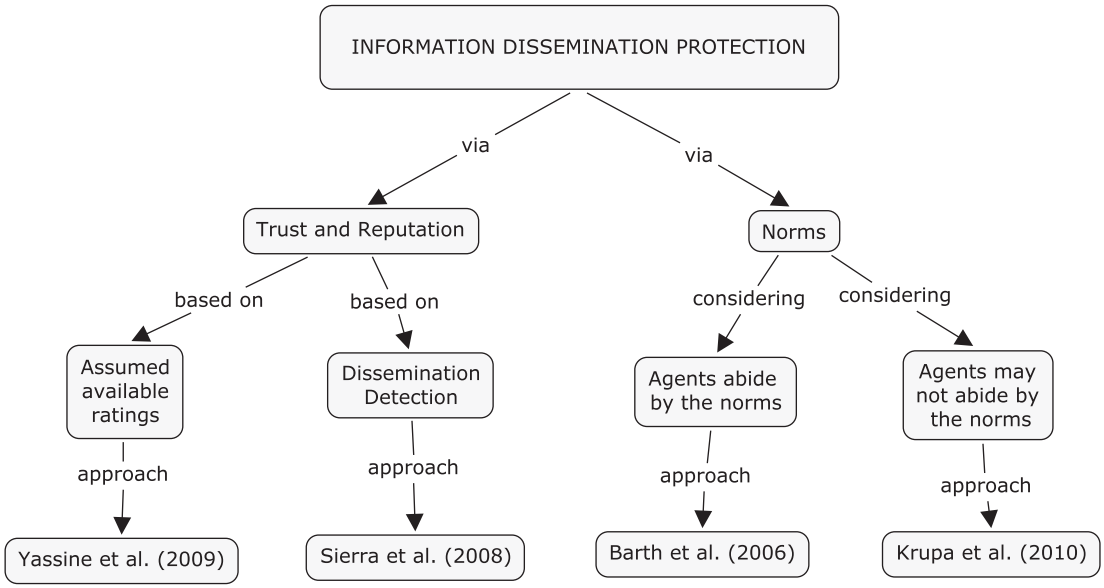


Figure 4 Information Dissemination Conceptual Map

In this way, we have described works that enhance privacy (by means of pseudonymity) but without impacting on security, trust, and reputation. Finally, we refer the reader to Section 6 to consult the open challenges on this.

5 Protection against information dissemination

Information dissemination refers to the transfer of previously collected and possibly processed data to other third parties. It should be pointed out that protection against dissemination in an open environment such as open multi-agent systems is a very hard problem. This is mainly because when a sender agent passes information to a receiver agent, the former usually loses control over that information. Moreover, it is very difficult for the sender agent to verify whether or not the receiver agent passes this information to other third parties. In the following, we outline some approaches to protect against information processing based on concepts usually used in agent-based technologies: trust and reputation, and norms.

Figure 4 depicts a conceptual map for all of the studied approaches that provide support for protecting against information dissemination.

5.1 Based on trust and reputation models

One approach to prevent information dissemination is based on trust and reputation models. There are works that assume that the reputation of another agent with regard to how they use the information they collect/process is available (Yassine & Shirmohammadi, 2009). Thus, agents can choose not to send information to agents that have a bad reputation. In this case, having a bad reputation means that the agent usually disseminates personal information about other agents. To measure the trustworthiness of agents regarding whether they disseminate personal information or not, one of the many trust and reputation models developed by the agent community could be used (refer to Ramchurn *et al.* (2004), and Sabater and Sierra (2005) for reviews on trust and reputation models).

These models usually need to verify the behavior of an agent in the past to predict their future behavior. However, how could an agent verify that another agent has disseminated information about it? The verification of whether or not an agent disseminates personal information about other agents or not is not straightforward. One approach could be that an external entity controls all the communications among agents. Thus, this external entity is able to know if an agent is

https://doi.org/10.1017/S0269888913000180 Published online by Cambridge University Press

disseminating information about another. This approach, however, cannot be applied due to its privacy implications (this external entity would act as a *big brother*). Instead, we envision multi-agent systems in which communications between each agent pair can be encrypted (by using mechanisms such as the ones presented in Section 3.2) to avoid undesired information collection by any other external entity.

Sierra and Debenham (2008) present a model for detecting undesired information dissemination based on information-theoretic measures. They consider that agents are uncertain about their world model. An agent estimates the amount of information that another agent possibly disseminated about the former agent from the information in the messages that the agent receives from other agents. To this aim, the agent sets update functions of its uncertain world model based on the messages received. For instance, if an agent A sends to agent B that A likes the color pink, agent A can set an update function of the messages received that scan for information related to the color pink. Thus, if an agent C sends agent A a message offering pink dresses, A could infer that B probably disseminated its color preferences to C. According to this, agent A can revisit the trustworthiness of agent B regarding information dissemination. This model only considers what an agent can observe by itself. However, other agents could also warn this agent about the fact that another agent is disseminating information about it. For instance, in the previous example, if agent A and agent C are known to each other, agent C may not take advantage of knowing A's color preferences. Instead, agent C can warn agent A that agent B disseminates information about it.

5.2 Based on normative multi-agent systems

This approach is based on using norms for governing the dissemination of information in a so-called normative multi-agent system (Criado *et al.*, 2011). According to Boella *et al.* (2007), a norm is 'a principle of right action binding upon the members of a group and serving to guide, control, or regulate proper and acceptable behavior'. In this case, proper and acceptable behavior means that agents should not be able to disseminate sensitive information about other agents without their consent.

Barth *et al.* (2006) present a logical framework for expressing and reasoning about norms of transmission of personal information. This framework formalizes the main ideas behind contextual integrity. Contextual integrity (Nissenbaum, 2004) is a legal framework for defining the appropriateness of the dissemination of information based on the context, the role of the entities taking part in the context, and the subject of the personal information being transferred. In the framework of Barth *et al.* (2006), privacy norms are expressed as linear temporal logic (LTL) formulas. These formulas are used to define the permissions and prohibitions when disseminating private information about other agents. For instance, it can be expressed that, in a medical context, an agent playing the role of doctor is allowed to pass medical personal information to an agent only if this agent is the subject of the information and this agent is playing the role of patient. Note the difference to an RBAC approach, which allows the definition of permission based on the roles of the entities taking part in the system. However, it cannot use the information about who is the subject of the information being transferred. Barth *et al.* (2006) assume a closed system in which all agents abide by the norms, so they do not provide any enforcement mechanism.

Krupa and Vercouter (2010) present a position paper that includes an initial proposal for controlling personal information dissemination in open multi-agent systems based on contextual integrity (explained above). They consider that agents may not abide by the norms. They propose five privacy-enforcing norms to promote privacy-preserving behaviors when disseminating information: (i) respect the appropriateness of the information to be transferred according to contextual integrity, that is, agents should not transfer information that is not appropriate regarding the context, the roles of the agents involved, and the subject of the information; (ii) sign the transmission chain before sending the information, so agents that transmit information remain liable for this transmission; (iii) do not send information to violating agents, that is, agents that do

https://doi.org/10.1017/S0269888913000180 Published online by Cambridge University Press

not abide by the norms; (iv) delete information received from violating agents so that this information is no longer transferred; and finally (v) punish agents violating these norms (including this one) by sending *punishment messages* (messages that inform that a given agent has performed a violation) to the subject of the information and also other agents in the system.

Krupa and Vercouter (2010) suggest the use of trust models based on the punishment messages to isolate violating agents, that is, if an agent is said to violate norms, other agents will not send personal information to it. Thus, trust and reputation models can be used based on these punishment messages. This is because, in this case, it is assumed that all of the agents will follow the norms, and in the event of not doing so, these punishment messages can act as the verification mechanism needed for trust and reputation models, as explained in Section 5.1. However, this work is at an initial stage. Specifically some major issues remain open (according to the authors): (i) the real connection of their proposal to trust models needs to be specified; (ii) two or more agents can easily collude by passing information to each other without other *benevolent* agents being aware of it; (iii) one agent can consider that another agent is not trustworthy according to its trust model, while another can consider it to be trustworthy (i.e. some transmissions can be viewed as appropriate by an agent and the same transmissions can be viewed as inappropriate by another agent); and (iv) the system can be subject to strategic manipulation, such as agents sending fake punishment messages that do not really correspond to real violations.

6 Open challenges

Over the course of this article, we have identified many possible lines for future research. In this section, we outline some of the most challenging possible future directions in the research field of privacy and MAS. These possible directions are open challenges identified during the realization of this survey. To our knowledge, all these open challenges will play a crucial role in broadening the use of agent technologies. On the one hand, principals will be more willing to engage with and delegate tasks to agents. On the other hand, agent technologies can be mixed/integrated with/into other information technologies to enhance the privacy that these information technologies provide.

6.1 Interoperability and openness

As stated by Luck *et al.* (2005), interoperability is crucial for the medium-term development of MAS. Interoperability is a basic requirement for building open MAS, in which heterogeneous agents can enter and leave the MAS at any moment, can interact with each other, and can span organizational boundaries. For instance, agent-based e-marketplaces are open MAS (Fasli, 2007a), in which buyer, seller, and broker agents can be developed by different developers using different languages and frameworks, so heterogeneity is inherent. Thus, agents and their interaction protocols need to allow interoperation. Thus, standards that help to allow interopertion are of crucial importance.

Although there are some standards proposed for agents and their interaction protocols, as yet there is no standard focusing on privacy issues. As described in Section 3.2, a first requirement for privacy is security. There have been some standards for security in MAS. FIPA defined a standard for security in MAS, but this standard soon became obsolete (FIPA, 1998). There have been some studies that consider this obsolete standard as a basis to analyze and propose guidelines for FIPA-based security standards (Poslad & Calisti, 2000; Poslad *et al.*, 2003). However, there has not been another proposal for a security standard for FIPA platforms since the obsolete standard from 1998.

The OMG Mobile Agent System Interoperability Facility (MASIF; Milojicic *et al.*, 1998) is a standard for mobile agents. Mobile agents are a species of autonomous agents that are capable of transmitting (migrating) themselves—their program and their state—across a computer network, and recommencing execution at a remote site (Wooldridge, 2002). MASIF specifies security mechanisms for mobile agents to migrate among hosts and also secure communication mechanisms.

https://doi.org/10.1017/S0269888913000180 Published online by Cambridge University Press

However, the security of MASIF is dependent on Corba-IDL mechanisms. No other transport mechanisms are considered, such as HTTP, AMQP, and others.

Security standards play a crucial role in preserving confidentiality in agent interactions. However, there are many other mechanisms that are needed for preserving privacy that also need to be standardized. For instance: how can agents selectively disclose parts of their identities in a standard way (as required by the disclosure decision-making mechanisms in Section 3.1)? How can agents change the pseudonyms they use in a standard way (as required for pseudonym management technologies in Section 4.2)? These standards do not need to be built from scratch. Instead, existing standards can be used as the basis for them. For instance, the OASIS Identity Metasystem Interoperability standard[10] is a standard for mechanisms that support pseudonymity and the selective disclosure of identity attributes.

6.2 Pseudonym changer agent

According to Hansen *et al.* (2008), one of the main questions that is relevant for pseudonyms to be privacy preserving is the amount of information that can be gathered by linking data disclosed under the same pseudonym. Social security numbers in the United States are a clear example that if a pseudonym is used for a long time, even spanning different contexts, different pieces of personal information disclosed in different contexts can be linked to each other, and also allow the inference of other personal information emerging from the combination of data in different contexts and applying learning and inference techniques. Moreover, linking a pseudonym to the real identity only once is sufficient to be able to associate all of this personal information to a real identity. This link can remain over time. Thus, other personal information disclosed under this pseudonym can be linked to the real identity of the principal subject of this personal information.

Clauβ *et al.* (2005) point out that different pseudonyms should be used in different contexts if the principal wants to maintain the personal information disclosed under each of these pseudonyms unlinkable. The most privacy-preserving option is to use transaction pseudonyms, which treat each transaction as a different context. However, there are many cases in which the principal can be interested in reusing the same pseudonym, for example, a social network that focus on a specific topic in which the principal is willing to establish friendships and other kinds of relationships that need the reuse of pseudonyms for recognizing entities from one time to another. Another example is that the principal itself could be willing to provide his/her profile to the seller agent in an e-commerce scenario in exchange for a discount or a reward, as pointed out in Section 3.1.

We mentioned several approaches for pseudonym support in Section 4.2. However, we could not find any approach that *suggests* pseudonym changes. In other words, the needed mechanisms for agents to be able to change their pseudonym exist, but there is no study or proposal for agents to decide when to change their pseudonym. This responsibility is given to the agent's designer or agent's principal. We envision pseudonym changer agents. These agents would be in charge of suggesting pseudonym changes by evaluating the privacy risks of reusing a pseudonym. Moreover, the models detailed in Section 3.1 could also be applied to make the decision of whether or not a pseudonym change is appropriate. This decision would take into account: the privacy risks due to not changing the pseudonym, and the utility or intimacy that would be lost by changing the pseudonym. Such *et al.* (2012b) propose the first step toward this ambitious endeavor.

6.3 Disclosure decision making based on multiple criteria

As stated in Section 3.1, current disclosure decision-making mechanisms are based on policies, the privacy-utility tradeoff, or social relationships. However, there is no proposal that brings these

[10] http://docs.oasis-open.org/imi/identity/v1.0/identity.html

https://doi.org/10.1017/S0269888913000180 Published online by Cambridge University Press

mechanisms together. This could be very appropriate for situations in which each one of these mechanisms is not enough by itself to cope with the requirements of agents' principals.

There are many examples of environments and situations in which these disclosure decision-making mechanisms can be combined. For instance, in a controlled environment in which policies are known to be enforced, the policies themselves can be used for agents that disclose personal information based on the privacy-utility tradeoff. In this way, agents are able to valuate the privacy loss they will suffer in the event of disclosing a specific piece of data, according to the policy of the intended destination agent. Then, based on this estimated privacy loss they can determine if the expected benefit for disclosing the information is worth it.

Another example could be the combination of the privacy-utility tradeoff with other more social approaches. For instance, suppose that an agent knows the benefit that disclosing personal information to another agent may cause to itself. Also suppose that this benefit is not worth the disclosure according to the privacy-utility tradeoff. An agent can still decide to disclose this information if it has a relationship that is intimate enough with the intended destination agent. Moreover, the agent could also decide to disclose this information if it does not have a relationship that is intimate enough with the intended destination agent, but it wants to reciprocate a previous disclosure of the intended destination agent.

Now, suppose that an agent has very low intimacy with another agent. Moreover, suppose that this low level of intimacy is due to the fact that the second agent deceived the first agent by not reciprocating its disclosures. The question arises as to whether or not the first agent should disclose personal information to the second agent if the first agent knows the utilitarian benefit of doing so and this benefit is high enough. In other words, how could the agent decide which of the two mechanisms to follow in a given situation?

6.4 Collective disclosure decision making

All of the disclosure decision-making mechanisms presented in Section 3.1 assume that the decision of whether or not to disclose is to be taken by only one agent. However, what to do when disclosing is a collective decision? There are many application domains in which the items to be disclosed may involve more than one agent, so that the privacy preferences of all of the agents involved should be satisfied when disclosing those items. For instance, in social networks there are items such as photos than may depict different people so that the privacy preferences of all of the people depicted should be respected (Gross & Acquisti, 2005). This can also occur in domains such as Virtual Organizations (Criado *et al.*, 2009) of agents that can produce items (that can contain sensitive information) and that must decide whether or not to disclose that items to other agents from outside the Virtual Organization. Finally, this can also occur generally in any of the existing collaborative software applications (Leuf & Cunningham, 2001) in which different parties must collectively decide whether or not to disclose and to whom to disclose the documents that are collaboratively authored.

A simple solution for this is to let agents to define their preferred privacy preferences with respect to an item and then merge these preferences into a disclosure policy for a given item. This merge would be valid as long as agent preferences do not conflict. However, there can be situations in which privacy preferences may conflict and this merge may not be obvious or may even be impossible. For these situations, we encourage research on methods and mechanisms to reach a common decision about whether or not to disclose and to whom to disclose items that involve more than one agent. These methods and mechanisms can be based on existing technologies to reach agreements among agents. These technologies are commonly referred to as Agreement Technologies (Sierra *et al.*, 2011), and involve technologies such as automated negotiation (Jennings *et al.*, 2001). Thus, agents could negotiate a common disclosure decision on a particular item.

6.5 Learning the privacy sensitivity of personal information

Most of the approaches presented in this survey assume that agents know the real privacy sensitivity of each personal attribute of their principals. However, this assumption is not always

https://doi.org/10.1017/S0269888913000180 Published online by Cambridge University Press

realistic. For instance, the number of personal attributes can be very large, so principals may not feel comfortable specifying the sensitivity for each of their personal attributes. Some of the approaches try to minimize this burden by clustering attributes into categories so that principals specify the sensitivity for each category (Yassine *et al.*, 2010). However, this can be also a burden if there is a huge number of categories or if the categories must be defined per target agent and there is a huge number of possible target agents.

A possible future line of research could be to automatically learn the privacy sensitivity of personal information based on studies such as the one presented by Huberman *et al.* (2005). They carried out an experiment that validates that people are usually more willing to disclose certain private attributes that are typical or positively atypical compared with the target group. The experiment assesses the value (in terms of monetary compensation) that people give to disclose personal attributes like weight and age. They gathered interesting results regarding weight (for age the significance was less): people who weigh less than the average required little compensation to disclose their weight, while people who weigh more than the average required a large compensation to disclose their weight. This is due to the fact that people who weigh more are afraid to feel embarrassment or stigmatization. The authors found a linear relationship about a trait and the value one places on it. The less desirable the trait, the more reluctant a person is to disclose the information. However, small deviations in a socially positive direction are associated with a lower monetary compensation request.

6.6 *Personal data attribute inference*

The decision-making models presented in Section 3.1 consider the privacy loss of disclosing a personal attribute before deciding whether they finally disclose it or not. This privacy loss usually considers the sensitivity of the personal attribute to be disclosed and the probability of linking this personal attribute to the real identity of the principal behind the agent. Although the agent can decide whether or not to disclose each attribute, it cannot control that other agents can infer other attributes that it does not want to disclose. This is known as the inference problem (Farkas & Jajodia, 2002). For instance, in the United States, if a principal discloses its driver license number, she/he is also disclosing that she/he is, at least, 16 years old.

Only a few of the decision-making models consider what could be inferred due to the disclosure of a personal attribute. Moreover, the decision-making models that consider these inferences provide very simple inference models. Several approaches tackle this problem in different computer science disciplines. These approaches are intended to infer the probabilities of linking personal data attributes to each other and to the principal they describe. For instance, there are approaches that deal with the inference problem when querying databases (Cuenca Grau & Horrocks, 2008), when applying data mining techniques (Zhu *et al.*, 2009), in social networks (Zheleva & Getoor, 2009), and in general, in all activities that require the publication of data (Chen *et al.*, 2009). All of these approaches consider complex models of personal information inference. The disclosure decision-making mechanisms for agents can either be based on these models or they can be adapted for the case of agents.

6.7 *Information dissemination detection*

As shown in Section 5, there are few studies that focus on information dissemination. Although these studies solve some of the problems that must be dealt with for protecting information dissemination, there are still many challenges that remain open. One such open challenge involves mechanisms for agents to detect when other agents disseminate information about them.

Sierra and Debenham (2008) propose an approach for an agent to detect that another agent is disseminating information about it based on scanning all the information the first agent receives in the search for clues of possible information disseminations. However, an agent may not be able to detect by itself that other agents are disseminating information about it. Another approach for information dissemination detection is based on notifications sent by other agents warning of

https://doi.org/10.1017/S0269888913000180 Published online by Cambridge University Press

possible dissemination of information. These notifications can play a crucial role when an agent itself is unable to detect that other agents are disseminating information about it. Krupa and Vercouter (2010) use notifications of disseminations in the form of what they call punishment messages. These messages are sent by the agent that detects an inappropriate dissemination to the rest of the agents, so that agents can know which agents perform inappropriate information disseminations. However, this mechanism can be subject to strategic manipulation, such as agents sending messages containing fake norm violations that do not really correspond to real violations.

6.8 Integration of trust, reputation, and norms for protecting against information dissemination

The real connection of norm-based approaches to trust-based approaches for avoiding information dissemination needs to be specified. This open challenge is also closely related to the information dissemination detection problem. If an agent is able to detect that another agent has performed information dissemination, it could revise the trust the first agent has in the second agent. Moreover, an agent could earn a very bad reputation with other agents by performing undesired information dissemination. In this way, both trust and reputation would act as privacy-enforcer mechanisms, isolating agents that disseminate information in an inappropriate way.

Krupa and Vercouter (2010) suggest that messages informing agents that violate the corresponding information dissemination norms can be used as inputs for trust and reputation models. Therefore, agents that do not abide by the norms would be considered as untrustworthy and would earn a bad reputation. This would finally result in the isolation of agents that do not abide by the norms. However, the authors of this work do not discuss how this integration of trust, reputation, and norms can be effectively achieved.

6.9 Avoiding collusion for protecting information dissemination

As shown in Section 5, current norm-based approaches to information dissemination are vulnerable to collusion. Thus, two or more agents could easily collude by passing information to each other without other benevolent and norm-abiding agents being aware of it. This could be addressed by using a central authority that would control and monitor the information that agents exchange. However, this may not be possible for various reasons. The main one, in line with this article, is privacy preservation. This is because this authority would become a *big brother*. Moreover, there may be other reasons, such as to prevent this authority from becoming a performance bottleneck and a SPOF.

Moreover, the problem of collusion could even be worse if we consider collusion in which one agent decides to disseminate information but without revealing the source of the information. Krupa and Vercouter (2010) identified this problem and called the agent that disseminates the information 'journalist'. As they state, a journalist agent would be an agent that decides to sacrifice himself to become a relay for information that violates the information dissemination norms. Therefore, the agent that is the source of the violation will never be known, and only the journalist will be seen as a violator of these norms. A journalist agent could even be rewarded with a monetary benefit in exchange for its practices.

6.10 Protection against information collection and dissemination

The combination/mixing of disclosure decision-making models with information dissemination models can play a crucial role preventing both information collection and information dissemination. All of the disclosure decision-making models presented in Section 3.1 assume separate interactions among agents for evaluating the privacy cost that a disclosure may cause. That is, these models do not consider that the agent that received the disclosure can share the received information with other agents.

An illustrative example can be: agent A decides by means of a disclosure decision-making models not to disclose its attribute location to agent B, for example, the expected utilitarian benefit

https://doi.org/10.1017/S0269888913000180 Published online by Cambridge University Press

for agent A to disclose its location to agent B is not high enough compared with the privacy loss this disclose may cause to agent A. However, in a different interaction, agent A decides, by means of a disclosure decision-making model, to disclose its attribute location to agent C, for example, the expected utilitarian benefit for agent A to disclose its location to agent C is high enough compared with the privacy loss. After this, agent A effectively discloses its location to agent C. Then, suppose that agent B and agent C are known to each other, so agent C can finally disclose the location of agent A to agent B. Therefore, B finally knows the location of agent A, even though agent A decided not to disclose it directly to agent B. Thus, we consider that if the dissemination risk is known, it should be considered when deciding whether to disclose information because if information collection is prevented, information dissemination cannot occur.

Another example can be: agent B has a bad reputation in a society because it usually disseminates the information it receives about other agents. Therefore, other agents in the society can decide to avoid disclosing information to B. However, suppose that an agent A and agent B have a very close relationship, that is, they have a medium/high degree of intimacy. Then, suppose that agent A has to decide whether to disclose its location to agent B. In this particular case, agent A could consider that the intimacy it has with agent B is high enough to assume that if it discloses new information to agent B, agent B will not disseminate it to other agents.

7 Conclusions

In this paper, we have introduced the issue of privacy preservation and its relation to multi-agent systems. We have also surveyed the state of the art on studies that fall into the intersection between privacy and multi-agent systems. We have classified them regarding the information-related activities that these studies aim to prevent: information collection, information processing, and information dissemination. First, there are works that prevent undesired information collection by means of disclosure decision- making mechanisms, by means of secure data transfer and storage infrastructures, and by means of trusted third parties that mediate communications. Second, there are works that prevent undesired information processing by using anonymity and pseudonymity techniques. Finally, there are works that prevent undesired information dissemination based on trust and reputation on the one hand and normative multi-agent systems on the other hand.

Although we have presented many studies that provide satisfactory solutions for some specific problems, we consider that research on privacy and multi-agent systems is still in its infancy. As pointed out in Section 6, there are still a great number of possible research lines that remain unexplored. To our knowledge, privacy will be a matter of major concern and will be the subject of many research efforts during this century. Multi-agent systems can play a crucial role for preserving privacy. To this aim, multi-agent systems need to preserve the privacy of the personal information that the agents that take part in these systems hold about real people (e.g. the personal information an agent could have about its principal in applications in which agents act on behalf of their principals). Agent-based solutions should also be integrated into other information technologies to enhance privacy. Thus, approaches based on concepts that are usually used in agent-based technologies (such as trust and reputation, and norms) can further enhance privacy in other existing information technologies.

Acknowledgments

This work has been partially supported by CONSOLIDER-INGENIO 2010 under grant CSD2007-00022, and project TIN2011-27652-C03-00 of the Spanish Government.

References

Ackerman, M. S., Cranor, L. F. & Reagle, J. 1999. Privacy in e-commerce: examining user scenarios and privacy preferences. In *EC'99: Proceedings of the 1st ACM Conference on Electronic Commerce*, Feldman, S. & Wellman, M. (eds). ACM, 1–8.

Acquisti, A., Gritzalis, S., Lambrinoudakis, C. & di Vimercati, S. (eds) 2008. *Digital Privacy: Theory, Technologies, and Practices*. Auerbach Publications.

https://doi.org/10.1017/S0269888913000180 Published online by Cambridge University Press

Aïmeur, E., Brassard, G., Fernandez, J. M. & Onana, F. S. M. 2006. Privacy-preserving demographic filtering. In *Proceedings of the 2006 ACM Symposium on Applied computing*, *SAC '06*, 872–878.

Alberola, J. M., Such, J. M., Garcia-Fornes, A., Espinosa, A. & Botti, V. 2010. A performance evaluation of three multiagent platforms. *Artificial Intelligence Review* **34**, 145–176.

Balke, T. & Eymann, T. 2008. The conclusion of contracts by software agents in the eyes of the law. In *Proceedings of the 7th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2008)*, 771–778.

Barth, A., Datta, A., Mitchell, J. C. & Nissenbaum, H. 2006. Privacy and contextual integrity: framework and applications. In *Proceedings of the 2006 IEEE Symposium on Security and Privacy*. IEEE Computer Society, 184–198.

Bhargav-Spantzel, A., Camenisch, J., Gross, T. & Sommer, D. 2007. User centricity: a taxonomy and open issues. *Journal of Computer Security* **15**, 493–527.

Bilge, L., Strufe, T., Balzarotti, D. & Kirda, E. 2009. All your contacts are belong to us: automated identity theft attacks on social networks. In *Proceedings of the 18th International Conference on World Wide Web*, *WWW '09*, Quemada, J., León, G., Maarek, Y. S. & Nejdl, W. (eds). ACM, 551–560.

Bishop, M. 2002. *Computer Security: Art and Science*. Addison-Wesley.

Boella, G., van der Torre, L. & Verhagen, H. 2007. Introduction to normative multiagent systems. In *Normative Multi-agent Systems, number 07122 in 'Dagstuhl Seminar Proceedings*, Boella, G., van der Torre, L. & Verhagen, H. (eds). Internationales Begegnungs- und Forschungszentrum für Informatik (IBFI).

Borking, J., Van Eck, B., Siepel, P. & Bedrijf, D. 1999. *Intelligent Software agents: Turning a Privacy Threat into a Privacy Protector*. Registratiekamer.

Bostwick, G. L. 1976. A taxonomy of privacy: repose sanctuary and intimate decisions. *California Law Review* **64**(5), 1447–1483.

Brazier, F., Oskamp, A., Prins, C., Schellekens, M. & Wijngaards, N. 2004. Anonymity and software agents: an interdisciplinary challenge. *Artificial Intelligence and Law* **12**, 137–157.

Brito, I. & Meseguer, P. 2003. Distributed forward checking. *Principles and Practice of Constraint Programming*, Lecture Notes in Computer Science **2833**, 801–806.

Bygrave, L. 2001. Electronic agents and privacy: a cyberspace odyssey 2001. *International Journal of Law and Information Technology* **9**(3), 275–294.

Camp, L. J. 1996. *Privacy & Reliability in Internet Commerce*. PhD thesis, Department of Computer Science, Carnegie Mellon University.

Carrara, E. & Hogben, G. 2007. Reputation-based systems: a security analysis, ENISA Position Paper.

Chaum, D. 1981. Untraceable electronic mail, return addresses, and digital pseudonyms. *Communications of the ACM* **24**, 84–90.

Chaum, D. 1985. Security without identification: transaction systems to make big brother obsolete. *Communications of the ACM* **28**, 1030–1044.

Chen, B.-C., Kifer, D., LeFevre, K. & Machanavajjhala, A. 2009. Privacy-preserving data publishing. *Foundations and Trends in Databases* **2**(1–2), 1–167.

Cheng, A. & Friedman, E. (2005), Sybil proof reputation mechanisms. In *Proceedings of the 2005 ACM SIGCOMM Workshop on Economics of Peer-to-Peer Systems*, *P2PECON '05*, Friedman, E. & Sirer, E. G. (eds). ACM, 128–132.

Chopra, S. & White, L. 2004. Artificial agents – personhood in law and philosophy. In *Proceedings of the 13th European Conference on Artificial Intelligence (ECAI 2004)*, 635–639.

Chopra, S. & White, L. 2007. Privacy and artificial agents, or, is google reading my email? In *Proceedings of the 20th International Joint Conference on Artificial Intelligence*, Sangal, R., Mehta, H. & Bagga, R. K. (eds). Morgan Kaufmann Publishers Inc., 1245–1250.

Cissèe, R. 2003. An architecture for agent-based privacy-preserving information filtering. In *Proceedings of the Sixth International Workshop on Trust, Privacy, Deception and Fraud in Agent Systems*.

Cissèe, R. & Albayrak, S. 2007. An agent-based approach for privacy-preserving recommender systems. In *Proceedings of the 6th International JointCconference on Autonomous Agents and Multiagent Systems*, *AAMAS '07*, Durfee, E., Yokoo, M., Huhns, M. & Shehory, O. (eds). ACM, 182:1–182:8.

Clauβ, S., Kesdogan, D. & Kölsch, T. 2005. Privacy enhancing identity management: protection against re-identification and profiling. In *DIM '05: Proceedings of the 2005 Workshop on Digital Identity Management*, Atluri, V., Samarati, P. & Goto, A. (eds). ACM, 84–93.

Crépin, L, Demazeau, Y., Boissier, O. & Jacquenet, F. 2009. Sensitive data transaction in hippocratic multi-agent systems. In *'Engineering Societies in the Agents World IX'*, Artikis, A., Picard, G. & Vercouter, L. (eds). Springer-Verlag, 85–101.

Crépin, L., Vercouter, L., Boissier, O., Demazeau, Y. & Jacquenet, F. 2008. Hippocratic multi-agent systems. In *Proceedings of the Tenth International Conference on Enterprise Information Systems ICEIS*, 301–307.

https://doi.org/10.1017/S0269888913000180 Published online by Cambridge University Press

Criado, N., Argente, E. & Botti, V. 2011. Open issues for normative multi-agent systems. *AI Communications* **24**(3), 233–264.

Criado, N., Argente, E., Julian, V. & Botti, V. 2009. Designing virtual organizations. In *Proceedings of the 7th International Conference on Practical Applications of Agents and Multi-Agent Systems (PAAMS2009), volume 55 of Advances in Soft Computing*, Salamanca, Spain, 440–449.

Cuenca Grau, B. & Horrocks, I. 2008. Privacy-preserving query answering in logic-based information systems. In *Proceedings of the 2008 Conference on ECAI 2008: 18th European Conference on Artificial Intelligence*, Ghallab, M., Spyropoulos, C., Fakotakis, N. & Avouris, N. (eds). IOS Press, 40–44.

Diaz, C. 2006. Anonymity metrics revisited. In *Anonymous Communication and its Applications, number 05411 in 'Dagstuhl Seminar Proceedings*, Dolev, S., Ostrovsky, R. & Pfitzmann, A. (eds). IBFI, 147–166.

Dierks, T. & Allen, C. 1999. The tls protocol version 1.0, RFC 2246, `http://www.ietf.org/rfc/rfc2246.txt` [retrieved February 2011].

Dingledine, R., Mathewson, N. & Syverson, P. 2004. Tor: the second-generation onion router. In *Proceedings of the 13th USENIX Security Symposium*, 303–320.

Farkas, C. & Jajodia, S. 2002. The inference problem: a survey. *ACM SIGKDD Explorations Newsletter* **4**(2), 11.

Fasli, M. 2007a. *Agent Technology for E-Commerce*. John Wiley & Sons.

Fasli, M. 2007b. On agent technology for e-commerce: trust, security and legal issues. *Knowledge Engineering Review* **22**(1), 3–35.

FIPA 1998. FIPA Agent Security Management, FIPA. `http://www.fipa.org/specs/fipa00020/OC00020A.html`

Fischer-Hübner, S. & Hedbom, H. 2008. Benefits of privacy-enhancing identity management. *Asia-Pacific Business Review* **10**(4), 36–52.

Friedman, E. J. & Resnick, P. 1998. The social cost of cheap pseudonyms. *Journal of Economics and Management Strategy* **10**, 173–199.

Frier, A., Karlton, P. & Kocher, P. 1996. The secure socket layer. Technical Report MSU-CSE-00-2, Netscape Communications.

Gangopadhyay, A. (ed.) 2001. *Managing Business with Electronic Commerce: Issues and Trends*. IGI Publishing.

Garfinkel, S. 2001. *Database Nation: The Death of Privacy in the 21st Century*. O'Reilly & Associates, Inc.

Garfinkel, S. 2009. Privacy requires security, not abstinence: protecting an inalienable right in the age of facebook, `http://www.technologyreview.com/computing/22831/`

Gibbins, N., Harris, S. & Shadbolt, N. 2004. Agent-based semantic web services. *Web Semantics: Science, Services and Agents on the World Wide Web* **1**(2), 141–154.

Goldschlag, D., Reed, M. & Syverson, P. 1999. Onion routing for anonymous and private internet connections. *Communications of the ACM* **42**, 39–41.

Green, K., Derlega, V. J. & Mathews, A. 2006. In *The Cambridge Handbook of Personal Relationships*. Cambridge University Press, chapter Self-Disclosure in Personal Relationships, 409–427.

Greenstadt, R., Grosz, B. & Smith, M. D. 2007. Ssdpop: improving the privacy of dcop with secret sharing. In *Proceedings of the 6th International Joint Conference on Autonomous Agents and Multiagent systems, AAMAS '07*, Durfee, E., Yokoo, M., Huhns, M. & Shehory, O. (eds). ACM, 171:1–171:3.

Greenstadt, R., Pearce, J. P. & Tambe, M. 2006. Analysis of privacy loss in distributed constraint optimization. In *Proceedings of the 21st National Conference on Artificial Intelligence – Volume 1*, Cohn, A. (ed.). AAAI Press, 647–653.

Gross, R. & Acquisti, A. 2005. Information revelation and privacy in online social networks. In *Proceedings of the 2005 ACM Workshop on Privacy in the Electronic Society*, Atluri, V., De Capitani di Vimercati, S. & Dingledine, R. (eds). ACM, 71–80.

Hansen, M., Berlich, P., Camenisch, J., Clau, S., Pfitzmann, A. & Waidner, M. 2004. Privacy enhancing identity management. *Information Security Technical Report* **9**(1), 35–44.

Hansen, M., Schwartz, A. & Cooper, A. 2008. Privacy and identity management. *IEEE Security & Privacy* **6**(2), 38–45.

Head, M. & Yuan, Y. 2001. Privacy protection in electronic commerce – a theoretical framework. *Human Systems Management* **20**(2), 149–160.

Hildebrandt, M. & Gutwirth, S. 2008. *Profiling the European Citizen: Cross-Disciplinary Perspectives*. Springer Publishing Company, Inc..

Hoffman, D., Novak, T. & Peralta, M. 1999. Building consumer trust online. *Communications of the ACM* **42**(4), 80–85.

Hoffman, K., Zage, D. & Nita-Rotaru, C. 2009. A survey of attack and defense techniques for reputation systems. *ACM Computing Surveys* **42**, 1:1–1:31.

Huberman, B. A., Adar, E. & Fine, L. R. 2005. Valuating privacy. *IEEE Security and Privacy* **3**(5), 22–25.

Ismail, L. 2008. A secure mobile agents platform. *Journal of Communications* **3**(2), 1–12.

https://doi.org/10.1017/S0269888913000180 Published online by Cambridge University Press

JADE Board 2005. Jade security guide, http://jade.tilab.com

Jaiswal, A., Kim, Y. & Gini, M. L. 2004. Design and implementation of a secure multi-agent marketplace. *Electronic Commerce Research and Applications* **3**(4), 355–368.

Jennings, N., Faratin, P., Lomuscio, A., Parsons, S., Wooldridge, M. & Sierra, C. 2001. Automated negotiation: prospects, methods and challenges. *Group Decision and Negotiation* **10**(2), 199–215.

Jøsang, A. & Golbeck, J. 2009. Challenges for robust trust and reputation systems. In *Proceedings of the 5th International Workshop on Security and Trust Management (STM 2009)*.

Jøsang, A., Ismail, R. & Boyd, C. 2007. A survey of trust and reputation systems for online service provision. *Decision Support System* **43**(2), 618–644.

Jøsang, A., Patton, M. & Ho, A. 2001. Authentication for humans. In *Proceedings of the 9th International Conference on; Telecommunication Systems (ICTS2001)*.

Kang, J. 1998. Information privacy in cyberspace transactions. *Stanford Law Review* **50**(4), 1193–1294.

Kerr, R. & Cohen, R. 2009. Smart cheaters do prosper: defeating trust and reputation systems. In *Proceedings of the 8th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2009)*, 993–1000.

Koops, B. -J. & Leenes, R. 2006. Identity theft, identity fraud and/or identity-related crime. *Datenschutz und Datensicherheit – DuD* **30**, 553–556.

Korba, L., Song, R. & Yee, G. 2002. Anonymous communications for mobile agents. In *Proceedings of the 4th International Workshop on Mobile Agents for Telecommunication Applications, MATA '02*, 171–181.

Krause, A. & Horvitz, E. 2008. A utility-theoretic approach to privacy and personalization. In *AAAI'08: Proceedings of the 23rd National Conference on Artificial Intelligence*, Cohn, A. (ed.). AAAI Press, 1181–1188.

Krupa, Y. & Vercouter, L. 2010. Contextual integrity and privacy enforcing norms for virtual communities. In *11th International Workshop on Coordination, Organization, Institutions and Norms in Multi-Agent Systems (COIN@MALLOW2010)*, 150–165.

Lebanon, G., Scannapieco, M., Fouad, M. R. & Bertino, E. 2006. Beyond k-anonymity: a decision theoretic framework for assessing privacy risk. In *Privacy in Statistical Databases*, 217–232.

Lee, H.-H. & Stamp, M. 2008. An agent-based privacy-enhancing model. *Information Management & Computer Security* **16**(3), 305–319.

Leuf, B. & Cunningham, W. 2001. *The Wiki Way: Quick Collaboration on the Web*. Addison-Wesley Longman Publishing Co., Inc.

Lilien, L. & Bhargava, B. 2008. Privacy and trust in online interactions. In *Online Consumer Protection: Theories of Human Relativism*, Chen, K. & Fadlalla, A. (eds). Information Science Reference, 85–122.

Longstaff, T., Ellis, J., Shawn, H., Lipson, H., Mcmillan, R., Pesante, H. L. & Simmel, D. 1997. Security of the internet. *The Froehlich/Kent Encyclopedia of Telecommunications* **15**, 231–255.

Luck, M., McBurney, P., Shehory, O. & Willmott, S. 2005. Agent Technology: Computing as Interaction (A Roadmap for Agent Based Computing). *AgentLink*.

Maheswaran, R., Pearce, J., Bowring, E., Varakantham, P. & Tambe, M. 2006. Privacy loss in distributed constraint reasoning: a quantitative framework for analysis and its applications. *Autonomous Agents and Multi-Agent Systems* **13**, 27–60.

Menczer, F., Street, W. N., Vishwakarma, N., Monge, A. E. & Jakobsson, M. 2002. Intellishopper: a proactive, personal, private shopping assistant. In *Proceedings of the First International Joint Conference on Autonomous Agents and Multiagent Systems: Part 3, AAMAS '02*, 1001–1008.

Milojicic, D., Breugst, M., Busse, I., Campbell, J., Covaci, S., Friedman, B., Kosaka, K., Lange, D., Ono, K., Oshima, M., Tham, C., Virdhagriswaran, S. & White, J. 1998. Masif: The omg mobile agent system interoperability facility. *Personal and Ubiquitous Computing* **2**, 117–129. 10.1007/BF01324942. http://dx.doi.org/10.1007/BF01324942

Mitchell, T., Caruana, R., Freitag, D., McDermott, J. & Zabowski, D. 1994. Experience with a learning personal assistant. *Communications of the ACM* **37**(7), 80–91.

Montaner, M., López, B. & De La Rosa, J. 2003. A taxonomy of recommender agents on the internet. *Artificial Intelligence Review* **19**(4), 285–330.

Newman, A. E. 2004. Cougaar developers' guide, http://www.cougaar.org

Neuman, C., Yu, T., Hartman, S. & Raeburn, K. 2005. *The Kerberos Network Authentication Service (V5), number 4120 in 'Request for Comments*. IETF.

Nissenbaum, H. 2004. Privacy as contextual integrity. *Washington Law Review* **79**(1).

Odlyzko, A. 2003. Privacy, economics, and price discrimination on the internet. In *Proceedings of the 5th International Conference on Electronic Commerce, ICEC '03*, Sadeh, N., Dively, M. J., Kauffman, M., Labrou, Y., Shehory, O., Telan, R. & Cranor, L. (eds). ACM, 355–366.

Olson, J. S., Grudin, J. & Horvitz, E. 2005. A study of preferences for sharing and privacy. In *CHI '05: CHI '05 Extended Abstracts on Human Factors in Computing Systems*, Van der Veer, G. & Gale, C. (eds). ACM, 1985–1988.

https://doi.org/10.1017/S0269888913000180 Published online by Cambridge University Press

Palen, L. & Dourish, P. 2003. Unpacking privacy for a networked world. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Ssystems*, Cockton, G. & Korhonen, P. (eds). ACM, 129–136.

Petkovic, M. & Jonker, W. (eds) 2007. *Security, Privacy and Trust in Modern Data Management (Data-Centric Systems and Applications)*. Springer-Verlag.

Pfitzmann, A. & Hansen, M. 2010. A terminology for talking about privacy by data minimization: anonymity, unlinkability, undetectability, unobservability, pseudonymity, and identity management, http://dud.inf.tu-dresden.de/Anon Terminology.shtml. v0.34.

Piolle, G., Demazeau, Y. & Caelen, J. 2007. Privacy management in user-centred multi-agent systems. In *Engineering Societies in the Agents World VII*, Lecture Notes in Computer Science **4457**, 354–367. Springer.

Poslad, S. & Calisti, M. 2000. Towards improved trust and security in FIPA agent platforms. In *Workshop on Deception, Fraud and Trust in Agent Societies*, Barcelona, Spain.

Poslad, S., Charlton, P. & Calisti, M. 2003. Specifying standard security mechanisms in multi-agent systems. In *Proceedings of the 2002 International Conference on Trust, Reputation, and Security: Theories and Practice, AAMAS'02*. Springer-Verlag, 163–176. http://portal.acm.org/citation.cfm?id=1762128.1762141

Quillinan, T. B., Warnier, M., Oey, M., Timmer, R. & Brazier, F. 2008. Enforcing security in the agent scape middleware. In *Proceedings of the 2008 Workshop on Middleware Security, MidSec '08*, Scandariato, R. & Russello, G. (eds). ACM, 25–30.

Ramchurn, S., Huynh, D. & Jennings, N. 2004. Trust in multi-agent systems. *The Knowledge Engineering Review* **19**(1), 1–25.

Rannenberg, K., Royer, D. & Deuker, A. (eds) 2009. *The Future of Identity in the Information Society: Challenges and Opportunities*. Springer Publishing Company, Incorporated.

Recursion Software Inc. 2008. Voyager security guide, http://www.recursionsw.com/

Roth, V. & Jalali-Sohi, M. 2001. Concepts and architecture of a security-centric mobile agent server. In *Proceedings of the International Symposium on Autonomous Decentralized System*.

Sabater, J. & Sierra, C. 2005. Review on computational trust and reputation models. *Artificial Intelligence Review* **24**, 33–60.

Schermer, B. 2007. *Software Agents, Surveillance, and the Right to Privacy: A Legislative Framework for Agent-enabled surveillance*. Amsterdam University Press.

Seigneur, J.-M. & Jensen, C. D. 2004. Trading privacy for trust. In *Trust Management*. Springer, 93–107.

Senicar, V., Jerman-Blazic, B. & Klobucar, T. 2003. Privacy-enhancing technologies–approaches and development. *Computer Standards & Interfaces* **25**(2), 147–158.

Shannon, C. E. 1948. A mathematical theory of communication. *Bell System Technical Journal* **27**(3), 379–423.

Sierra, C. 2004. Agent-mediated electronic commerce. *Autonomous Agents and Multi-Agent Systems* **9**(3), 285–301.

Sierra, C., Botti, V. & Ossowski, S. 2011. Agreement computing. *KI-Künstliche Intelligenz* **25**(1), 57–61.

Sierra, C. & Debenham, J. 2008. Information-based deliberation. In *AAMAS '08: Proceedings of the 7th International Joint Conference on Autonomous Agents and Multiagent Systems*, Padgham, L., Parkes, D. C., Müller, J. P. & Parsons, S. (eds). International Foundation for Autonomous Agents and Multiagent Systems, 689–696.

Silaghi, M. C. & Mitra, D. 2004. Distributed constraint satisfaction and optimization with privacy enforcement. In *Proceedings of the IEEE/WIC/ACM International Conference on Intelligent Agent Technology*, *IAT '04*. IEEE Computer Society, 531–535.

Smith, H. J. & Milberg, S. J. 1996. Information privacy: measuring individuals' concerns about organizational practices. *MIS Quarterly* **20**, 167–196.

Solanas, A. & Martìnez-ballesté, A. 2009. *Advances in Artificial Intelligence for Privacy Protection and Security*. World Scientific Publishing Co., Inc.

Solove, D. 2002. Conceptualizing privacy. *California Law Review* **90**(4), 1087–1155.

Solove, D. 2006. A taxonomy of privacy. *University of Pennsylvania Law Review* **154**(3), 477–560.

Spiekermann, S. 2006. Individual price discriminaton – an impossibility?. In *International Conference for Human–Computer Interaction (CHI'2006)*, Workshop on Privacy and Personalization'.

Spiekermann, S. & Cranor, L. F. 2009. Engineering privacy. *IEEE Transactions on Software Engineering* **35**(1), 67–82.

Stallings, W. 2010. *Network Security Essentials: Applications and Standards*. Prentice Hall.

Stamp, M. 2006. *Information Security: Principles and Practice*. Wiley-Interscience.

Such, J. M. 2011. *Enhancing Privacy in Multi-agent Systems*. PhD thesis, Departament de Sistemes Informàtics i Computació, Universitat Politècnica de València.

Such, J. M., Alberola, J. M., Espinosa, A. & Garcia-Fornes, A. 2011a. A group-oriented secure multiagent platform. *Software: Practice and Experience* **41**(11), 1289–1302.

Such, J. M., Garcia-Fornes, A., Espinosa, A. & Bellver, J. 2013. Magentix2: A privacy-enhancing Agent Platform. *Engineering Applications of Artificial Intelligence* **26**(1), 96–109.

https://doi.org/10.1017/S0269888913000180 Published online by Cambridge University Press

Such, J. M., Espinosa, A., Garcia-Fornes, A. & Botti, V. 2011b. Partial identities as a foundation for trust and reputation. *Engineering Applications of Artificial Intelligence* **24**(7), 1128–1136.

Such, J. M., Espinosa, A., Garcia-Fornes, A. & Sierra, C. 2011d. Privacy-intimacy tradeoff in selfdisclosure. In *Proceedings of the 10th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS2011)*. IFAAMAS, 1189–1190.

Such, J. M., Espinosa, A., Garcia-Fornes, A. & Sierra, C. 2012a. Self-disclosure decision making based on intimacy and privacy. *Information Sciences* **211**, 93–111.

Such, J. M., Serrano, E., Botti, V. & Garca-Fornes, A. 2012b. Strategic pseudonym change in agent-based E-commerce. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2012)*, Valencia, Spain, 1377–1378.

Taylor, H. 2003. Most People Are "Privacy Pragmatists" Who, While Concerned about Privacy, Will Sometimes Trade It Off for Other Benefits., Harris Interactive. Retrieved February 27, 2011, from `http://www.harrisinteractive.com/vault/Harris-Interactive-Poll-Research-Most-People-Are-Privacy-Pragmatists-Who-While-Conc-2003-03.pdf`

Tentori, M., Favela, J. & Rodriguez, M. D. 2006. Privacy-aware autonomous agents for pervasive healthcare. *IEEE Intelligent Systems* **21**, 55–62.

Udupi, Y. B. & Singh, M. P. 2010. Information sharing among autonomous agents in referral networks. In *Agents and Peer-to-Peer Computing*, Joseph, S. R., Despotovic, Z., Moro, G. & Bergamaschi, S. (eds). Springer-Verlag, 13–26.

Ugurlu, S. & Erdogan, N. 2005. An overview of secmap secure mobile agent platform. In *Proceedings of Second International Workshop on Safety and Security in Multiagent Systems*.

van Blarkom, G., Borking, J. & Olk, J. (eds) 2003. *Handbook of Privacy and Privacy-Enhancing Technologies: The Case of Intelligent Software Agents*. College bescherming persoonsgegevens.

van Elst, L., Abecker, A. & Dignum, V. 2004. *Agent-mediated Knowledge Management: International Symposium Amkm 2003, Stanford, CA, USA, March 24–26, 2003: Revised and Invited Papers*, Lecture Notes in Computer Science **2926**, 428. Springer-Verlag.

van Otterloo, S. 2005. The value of privacy: optimal strategies for privacy minded agents. In *AAMAS '05: Proceedings of the Fourth International Joint conference on Autonomous Agents and Multiagent Systems*, Pechoucek, M., Steiner, D. & Thompson, S. (eds). ACM, 1015–1022.

Wallace, R. J. & Freuder, E. C. 2005. Constraint-based reasoning and privacy/efficiency tradeoffs in multi-agent problem solving. *Artificial Intelligence* **161**, 209–227.

Warnier, M. & Brazier, F. 2010. Anonymity services for multi-agent systems. *Web Intelligence and Agent Systems* **8**(2), 219–232.

Warren, S. & Brandeis, L. 1890. The right to privacy. *Harvard Law Review* **4**(5), 193–220.

Weitnauer, E., Thomas, N., Rabe, F. & Kopp, S. 2008. Intelligent agents living in social virtual environments bringing max into second life. In *Intelligent Virtual Agents (IVA)*. Springer, 552–553.

Westin, A. 1967. *Privacy and Freedom*. Atheneum.

Westin, A. 1984. The origins of modern claims to privacy. In *Philosophical Dimensions of Privacy: An Anthology*, Schoeman, F. D. (ed.). Cambridge University Press, 56–74.

Wooldridge, M. 2002. *An Introduction to MultiAgent Systems*. Wiley.

Xu, H. & Shatz, S. M. 2003. Adk: an agent development kit based on a formal design model for multi-agent systems. *Journal of Automated Software Engineering* **10**, 337–365.

Yao, M. Z., Rice, R. E. & Wallis, K. 2007. Predicting user concerns about online privacy. *J American Society for Information Science and Technology* **58**(5), 710–722.

Yassine, A., Shirehjini, A. A. N., Shirmohammadi, S. & Tran, T. T. 2010. An intelligent agent-based model for future personal information markets. *IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, **2**, 457–460, Toronto, Canada.

Yassine, A. & Shirmohammadi, S. 2009. Measuring users' privacy payoff using intelligent agents. In *Computational Intelligence for Measurement Systems and Applications*, *CIMSA '09*, 169–174.

Yokoo, M., Suzuki, K. & Hirayama, K. 2005. Secure distributed constraint satisfaction: reaching agreement without revealing private information. *Artificial Intelligence* **161**, 229–245.

Yu, H., Kaminsky, M., Gibbons, P. B. & Flaxman, A. 2006. Sybilguard: defending against sybil attacks via social networks. In *Proceedings of the 2006 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications*, *SIGCOMM '06*, Rizzo, L., Anderson, T. & McKeown, N. (eds). ACM, 267–278. `http://doi.acm.org/10.1145/1159913.1159945`

Zheleva, E. & Getoor, L. 2009. To join or not to join: the illusion of privacy in social networks with mixed public and private user profiles. In *WWW '09: Proceedings of the 18th International Conference on World Wide Web*, Quemada, J., León, G., Maarek, Y. S. & Nejdl, W. (eds). ACM, 531–540.

Zhu, Z., Wang, G. & Du, W. 2009. Deriving private information from association rule mining results. In *Proceedings of the 2009 IEEE International Conference on Data Engineering*, Ioannidis, Y. E., Lee, D. L. & Ng, R. T. (eds). IEEE Computer Society, 18–29.

https://doi.org/10.1017/S0269888913000180 Published online by Cambridge University Press